



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/87, 15/62, C07K 14/395, 14/55, A61K 48/00		A1	(11) International Publication Number: WO 95/28494
			(43) International Publication Date: 26 October 1995 (26.10.95)
(21) International Application Number: PCT/US95/04738		(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).	
(22) International Filing Date: 17 April 1995 (17.04.95)		<p>Published</p> <p><i>With international search report.</i></p> <p><i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	
(30) Priority Data:			
08/227,858	15 April 1994 (15.04.94) US		
08/326,460	19 October 1994 (19.10.94) US		
(60) Parent Application or Grant			
(63) Related by Continuation			
US	08/326,460 (CIP)		
Filed on	19 October 1994 (19.10.94)		
(71) Applicant (for all designated States except US): TARGETED GENETICS CORPORATION [US/US]; Suite 100, 1100 Olive Way, Seattle, WA 98101 (US).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): OVERELL, Robert, W. [US/US]; 3209 151st Place, S.E., Mill Creek, WA 98012 (US). WEISSER, Karen, E. [US/US]; 3413 19th Avenue South, Seattle, WA 98144 (US).			
(74) Agents: DYLAN, Tyler et al.; Morrison & Foerster, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).			
(54) Title: GENE DELIVERY FUSION PROTEINS			
(57) Abstract			
<p>The invention provides gene delivery fusion protein (GDFPs) for use in gene transduction of target cells, such as mammalian cells. The GDFP contains a nucleic acid binding domain (NBD) that binds to a targeted nucleic acid to be transduced, fused to a gene delivery domain (GDD) that mediates or augments transfer of the targeted nucleic acid into the target cell. The GDD contains one or more components that facilitate gene delivery, including binding/targeting components, membrane-disrupting components, transport/localization components and replicon integration components.</p>			
<p>GDFP</p> <p>FLEXION</p> <p>GDD</p> <p>NBD</p> <p>TARGETED NUCLEIC ACID (INA)</p> <p>COGNATE RECOGNITION SEQUENCE</p> <p>NBD BINDS GDFP TO INA</p> <p>GDD BINDS INA-GDFP COMPLEX TO TARGET CELL AND MEDIATES GENE DELIVERY.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

GENE DELIVERY FUSION PROTEINS

Technical Field

5 The invention relates to the field of gene delivery, more specifically to proteins useful for introducing polynucleotides into target cells. Still more specifically, the invention relates to fusion proteins that are capable of both binding to a polynucleotide of interest, and of facilitating delivery of the bound polynucleotide to a target cell, especially to a mammalian target cell.

10

Background

Many viruses have been adapted for use as gene delivery vectors for mammalian cells. Viruses have highly efficient mechanisms for entering cells, and in some cases also have specific mechanisms for integrating the viral genome into the host cell
15 chromosome. The high efficiency of gene transduction afforded by the viral vectors is the principal advantage of using a virus-based system for gene delivery. In addition, the fact that the viruses are particulate allows virus-based systems to be considered for in vivo gene delivery. These attributes have led to the wide use of viral vectors in gene transfer studies. Viruses that have been used for this purpose include
20 retroviruses, adenoviruses, parvoviruses, papovaviruses, poxviruses and herpesviruses. More recently, the utility of viral vectors has led to the use of retroviruses and adenoviruses in gene therapy applications.

Although the virus-based delivery systems can give rise to high efficiency of gene delivery, they suffer from a number of disadvantages. For example, the most
25 widely used viral system, the retroviral vectors, have been extensively modified to prevent the generation of replication-competent retrovirus (RCR), but since such RCR has the potential to be leukemogenic (see Donahue et al., J. Exp. Med. 176:1125-1135,

-2-

1992), all retroviral preparations for use in gene therapy must undergo extensive validation testing to confirm the absence of RCR before use. In addition to these safety concerns, retroviral and other viral vectors can place size and sequence constraints on the genetic material that can be transferred and on the target cells that can be infected
5 (see, e.g., Israel & Kaufman, *Blood*, 75:1074-1080, 1990; Shimotohno & Temin, *Nature* 299,265-268, 1982; Stead et al., *Blood*, 71:742-747, 1988; and Bodine et al., *Blood*, 82:1975-1980, 1993).

The development of efficient non-viral gene delivery (NVGD) systems would allow gene transfer/gene therapy studies to be performed in the absence of the
10 aforementioned limitations of the viral vectors, and could also have the advantages of ease of scalability, cost and speed of generation. Based on these advantages, non-viral gene delivery systems could also allow more diseases to be treated through gene therapy by making injectable gene delivery systems a reality.

Existing non-viral gene delivery systems can be roughly divided into physical
15 and biochemical approaches. The physical methods include such techniques as electroporation, particle bombardment, scrape loading and calcium phosphate transfection (see, e.g., Fechheimer et al., *P.N.A.S.* 84:8463-8467, 1987; Cheng et al., *P.N.A.S.* 90:4455-4459, 1993; and Kriegler, M. (ed.), "Gene Transfer and Expression, a Laboratory Manual," 1990, W.H. Freeman Publishers). The
20 biochemical methods involve mixing the DNA to be delivered with reagents such as DEAE-dextran, gramicidin S, liposomes, polyamidoamine polymers, polyamines, polybrene, cationic proteins and poly-L-lysine-based conjugates (see, e.g., Kawai & Nishizawa, *Mol. Cell. Biol.* 4:1172-1174, 1984; Behr et al., *P.N.A.S.* 86:6982-6986, 1989; Rose et al., *P.N.A.S. Biotechniques* 10:520-525, 1991; Pardridge & Boado,
25 *F.E.B.S. Lett.* 288:30-32, 1991; Legendre & Szoka, *P.N.A.S.* 90:893-897, 1993; Haensler & Szoka, *Bioconj. Chem.* 4:372-379, 1993; and Wu and Wu, *J. Biol. Chem.* 262:4429-4432, 1987).

These different approaches vary in their efficiency of gene delivery and in their ability to confer long-term (i.e. stable) retention of transferred sequences. However,
30 the biochemical approaches are in general more attractive from a gene therapy point of view because such approaches have a greater potential for use within injectable gene delivery systems than do most of the physical approaches.

-3-

One problem with the use of conjugates based on poly-L-lysine or other basic polymers, which are assembled via chemical cross-linking, is that the chemical steps required for cross-linking can be both imprecise and cumbersome. Moreover, it can be very difficult to control the stoichiometry of the different conjugate components in such a system, particularly as more components are added to facilitate gene delivery.

Summary of the Invention

In view of the continuing and unmet need for safe, efficient and stable non-viral gene delivery systems, the present invention provides a generalized approach for the modular construction of fusion proteins that are capable of both binding to a polynucleotide of interest, and of facilitating delivery of the bound polynucleotide to a target cell, especially to a human target cell for gene therapy.

The proteins of the present invention, termed Gene Delivery Fusion Proteins (GDFPs) comprise a nucleic acid binding domain (NBD) that contains a component capable of binding the targeted nucleic acid; fused to a gene delivery domain (GDD) that contains one or more components that mediate or facilitate delivery of the targeted nucleic acid to the target cell.

As described in detail below, nucleic acid binding domains can comprise any of a number of components, the essential feature of which is that they are capable of binding nucleic acids. A number of such components are known in the art (see, e.g., the references cited below), including proteins that bind nucleic acids in a sequence-specific manner and proteins that bind nucleic acids relatively non-specifically. For purposes of discussion and illustration, nucleic acid binding domains can be conveniently grouped into either of two basic subsets depending on whether the nucleic acid binding domain does or does not contain an analog of a sequence-specific nucleic acid binding protein, as described in more detail below.

In a first type of gene delivery fusion protein of the present invention (sometimes referred to herein as a "Type-I GDFP"), the nucleic acid binding domain contains an analog of a sequence-specific nucleic acid binding protein (sequence-specific NBP). In a second type of gene delivery fusion protein of the present invention (sometimes referred to herein as a "Type-II GDFP"), the nucleic acid binding domain contains an analog of a sequence-non-specific nucleic acid binding protein

(sequence-non-specific NBP) and does not contain an analog of a sequence-specific NBP.

Thus, one embodiment of a GDFP of the present invention is a macromolecule useful in delivering a targeted nucleic acid to a target cell, comprising a gene delivery fusion protein (GDFP), said GDFP comprising a nucleic acid binding domain (NBD) that contains a component capable of binding to a cognate recognition sequence in the targeted nucleic acid which component is derived from a sequence-specific nucleic acid binding protein; fused to a gene delivery domain (GDD) that contains one or more components that mediate or facilitate delivery of the targeted nucleic acid to the target cell. In addition to the binding component derived from a sequence-specific NBP, the nucleic acid binding domain of Type-I GDFPs can also contain additional binding components, as discussed below, which can be derived from either sequence-specific or sequence-non-specific nucleic acid binding proteins.

Another embodiment of a GDFP of the present invention is a macromolecule useful in delivering a targeted nucleic acid to a target cell, comprising a gene delivery fusion protein (GDFP), said GDFP comprising a nucleic acid binding domain (NBD) that contains a component capable of binding the targeted nucleic acid which component is an analog of a sequence-non-specific nucleic acid binding protein; fused to a gene delivery domain (GDD) that contains one or more components that mediate or facilitate delivery of the targeted nucleic acid to the target cell.

In one aspect of the invention, the components of the gene delivery domain (GDD) that facilitate delivery of the targeted nucleic acid to the target cell are selected from the group consisting of a binding/targeting component, a membrane-disrupting component, a transport/localization component and a replicon integration component. In another aspect of the invention, the various functional domains and components of the GDFP are separated by flexible peptide linker sequences ("flexons"), which can enhance the ability of the components to adopt conformations relatively independently of each other.

Another embodiment of a GDFP of the invention is a recombinant polynucleotide encoding a GDFP. In a preferred embodiment of this type, the polynucleotide is an expression vector and is arranged so that the various domains and components of the GDFP are expressed as an in-frame fusion product, thereby allowing

-5-

for efficient modular synthesis of the GDFP as a single recombinant product. Yet another embodiment is a method of using the above-described recombinant polynucleotide to produce a GDFP, said method comprising the steps of causing the recombinant polynucleotide to be transcribed and/or translated and recovering a GDFP.

5 As discussed herein, the preferred method involves the modular synthesis of the GDFP as a single protein product. Yet another embodiment is a method of using a GDFP to deliver a targeted nucleic acid (tNA) to a target cell, the method comprising the steps of contacting the GDFP with the targeted nucleic acid to produce a GDFP/nucleic acid complex and contacting said GDFP/nucleic acid complex with the target cell.

10 Preferably, the tNA is an expression vector. Preferably, the target cell is a mammalian cell. Yet another embodiment is a cell produced by the above-described method of using a GDFP and the progeny thereof.

15

Brief Description of the Drawings

Figure 1 is a schematic representation of an embodiment of the Gene Delivery Fusion Protein (GDFP) concept using a Type-I GDFP.

Figures 2A and 2B are diagrams of the cloning strategy used to generate expression vectors encoding IL-2, GAL4, and the GAL4/IL-2 and IL-2/GAL4 GDFPs.

20

Figure 3 is an SDS-PAGE gel of 35-S labeled GAL4/IL-2m GDFP.

Figure 4 is a gel-shift assay showing retention of DNA binding activity by the GAL4/IL-2m GDFP.

Figure 5 shows retention of IL-2 bioactivity by the GAL4/IL-2 GDFP.

Figure 6 is an SDS-PAGE gel of 35-S labeled GAL4/IL-2 GDFP and IL-2/GAL4 GDFP.

25

Figure 7 shows sequence-specific DNA binding of the GAL4 protein and the IL-2/GAL4 and GAL4/IL-2 GDFPs.

Figure 8 shows the cytokine bioactivity of the IL-2/GAL4 and GAL4/IL-2 GDFPs.

30

Figure 9 shows the results of an assay demonstrating the ability of GDFPs to bind to IL-2 receptor-bearing CTLL.

-6-

Figure 10 shows the results of an assay demonstrating the ability of GAL4/IL-2 GDFP and IL-2/GAL4 GDFP to mediate binding of a target oligomer to IL-2 receptor-bearing CTLL.

Figure 11 shows the results of an assay demonstrating the ability of GAL4/IL-2 GDFP to mediate binding of a target plasmid to IL-2 receptor-bearing CTLL.

Detailed Description of the Invention

The invention provides a non-viral gene delivery system by which DNA, RNA and/or analogs thereof ("targeted nucleic acid" or "tNA" to be used in gene delivery) are modified by association with a gene delivery fusion protein (GDFP). The non-viral gene delivery system of the present invention comprises a macromolecular complex of two separate entities: the targeted nucleic acid to be delivered, and a GDFP. The GDFP comprises a nucleic acid binding domain (NBD) that can bind to the targeted nucleic acid and thus lead to the formation of a GDFP/tNA complex; fused to a gene delivery domain (GDD) that can mediate or facilitate the delivery of the GDFP/tNA complex into the target cells.

In a preferred embodiment of the invention the open reading frames encoding the various GDFP domains and components are fused to enable expression of the GDFP as a single polypeptide. However, the GDFP may also comprise, for example, one or more short flexible peptide linker sequence ("flexons") between the individual domains and/or components.

General Definitions

The terms "polypeptide", "peptide" and "protein" are used interchangeably to refer to polymers of amino acids and do not refer to any particular lengths of the polymers. These terms also include post-translationally modified proteins, for example, glycosylated, acetylated, phosphorylated proteins and the like. Also included within the definition are, for example, proteins containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), proteins with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

-7-

"Native" polypeptides or polynucleotides refer to polypeptides or polynucleotides recovered from a source occurring in nature. Thus, the phrase "native viral binding proteins" would refer to naturally occurring viral binding proteins.

5 "Mutein" forms of a protein or polypeptide are those which have minor alterations in amino acid sequence caused, for example, by site-specific mutagenesis or other manipulations; by errors in transcription or translation; or which are prepared synthetically by rational design. Minor alterations are those which result in amino acid sequences wherein the biological activity of the polypeptide is retained and/or wherein the mutein polypeptide has at least 90% homology with the native form.

10 An "analog" of a polypeptide X includes fragments and muteins of polypeptide X that retain a particular biological activity; as well as polypeptide X that has been incorporated into a larger molecule (other than a molecule within which it is normally found); as well as synthetic analogs that have been prepared by rational design. For example, an analog of a DNA binding protein might refer to a portion of a native DNA
15 binding protein that retains the ability to bind to DNA, to a mutein thereof, to an entire native binding protein that has been incorporated into a recombinant fusion protein, or to an analog of a native binding protein that has been synthetically prepared by rational design.

"Polynucleotide" refers to a polymeric form of nucleotides of any length, either
20 ribonucleotides or deoxyribonucleotides, or analogs thereof. This term refers only to the primary structure of the molecule. Thus, double- and single-stranded DNA, as well as double- and single- stranded RNA are included. It also includes modified polynucleotides such as methylated or capped polynucleotides.

An "analog" of DNA, RNA or a polynucleotide, refers to a macromolecule
25 resembling naturally-occurring polynucleotides in form and/or function (particularly in the ability to engage in sequence-specific hydrogen bonding to base pairs on a complementary polynucleotide sequence) but which differs from DNA or RNA in, for example, the possession of an unusual or non-natural base or an altered backbone. A large variety of such molecules have been described for use in antisense technology;
30 see, e.g., E. Uhlmann et al. (1990) Chemical Reviews 90:543-584, and the publications reviewed therein.

-8-

An "antisense" copy of a particular polynucleotide refers to complementary sequence that is capable of hydrogen bonding to the polynucleotide and may, therefore, be capable of modulating expression of the polynucleotide (i.e. by "antisense" regulation). Such an antisense copy may be DNA, RNA or analogs thereof, including
5 analogs having altered backbones, as described above. The polynucleotide to which the antisense copy binds may be in single-stranded form (such as an mRNA molecule) or in double-stranded form (such as a portion of a chromosome).

A "replicon" refers to a polynucleotide comprising an origin of replication (generally referred to as an ori sequence) which allows for replication of the
10 polynucleotide in an appropriate host cell. Examples include replicons of a target cell into which a desired nucleic acid might integrate (in particular, nuclear and mitochondrial chromosomes; and also extrachromosomal replicons such as plasmids).

"Recombinant," as applied to a polynucleotide, means that the polynucleotide is the product of various combinations of cloning, restriction and/or ligation steps
15 resulting in a construct that is distinct from a polynucleotide found in nature. "Recombinant" may also be used to refer to the protein product of a recombinant polynucleotide. Typically, DNA sequences encoding the structural coding sequence for, e.g., components of the NBD and GDD, can be assembled from cDNA fragments and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a
20 synthetic gene which is capable of being expressed when operably linked to a transcriptional regulatory region. Such sequences are preferably provided in the form of an open reading frame uninterrupted by internal non-translated sequences (i.e. "introns"), such as those commonly found in eukaryotic genes. Such sequences, and all of the sequences referred to in the context of the present invention, can also be
25 generally obtained by PCR amplification using viral, prokaryotic or eukaryotic DNA or RNA templates in conjunction with appropriate PCR amplimers.

A "recombinant expression vector" refers to a polynucleotide which contains a transcriptional regulatory region and coding sequences necessary for the expression of an RNA molecule and/or protein and which is capable of being introduced into a target
30 cell (by, e.g., viral infection, transfection, electroporation or by the non-viral gene delivery (NVGD) techniques of the present invention). A further example would be an expression vector used to express a GDFP of the present invention.

-9-

"Recombinant host cells", "host cells", "cells", "target cells", "cell lines", "cell cultures", and other such terms denote higher eukaryotic cells, most preferably mammalian cells, which can be, or have been, used as recipients for recombinant vectors or other transfer polynucleotides, and include the progeny of the original cell which has been transduced. It is understood that the progeny of a single cell may not necessarily be completely identical (in morphology or in genomic or total DNA complement) to the original parent cell, due to natural, accidental, or deliberate mutation.

An "open reading frame" (or "ORF") is a region of a polynucleotide sequence that can encode a polypeptide or a portion of a polypeptide (i.e., the region may represent a portion of a protein coding sequence or an entire protein coding sequence).

"Fused" or "fusion" refers to the joining together of two or more elements, components, etc., by whatever means (including, for example, a "fusion protein" made by chemical conjugation (whether covalent or non-covalent), as well as the use of an in-frame fusion to generate a "fusion protein" by recombinant means, as discussed infra).

An "in-frame fusion" refers to the joining of two or more open reading frames (ORFs), by recombinant means, to form a single larger ORF, in a manner that maintains the correct reading frame of the original ORFs. Thus, the resulting recombinant fusion protein is a single protein containing two or more segments that correspond to polypeptides encoded by the original ORFs (which segments are not normally so joined in nature). Although the reading frame is thus made continuous throughout the fused segments, the segments may be physically separated by, for example, in-frame flexible polypeptide linker sequences ("flexons"), as described infra.

A "flexon" refers to a flexible polypeptide linker sequence (or to a nucleic acid sequence encoding such a polypeptide) which typically comprises amino acids having small side chains (e.g., glycine, alanine, valine, leucine, isoleucine and serine). In the present invention, flexons can be incorporated in the GDFP between one or more of the various domains and components. Incorporating flexons between these components is believed to promote functionality by allowing them to adopt conformations relatively independently from each other. Most of the amino acids incorporated into the flexon will preferably be amino acids having small side chains. The flexon will preferably comprise between about four and one hundred amino acids, more preferably between

-10-

about eight and fifty amino acids, and most preferably between about ten and thirty amino acids. Flexon ("Pixy") sequences described in U.S. Patents 5,073,627 and 5,108,910 will also be suitable for use as flexons.

5 A "transcriptional regulatory region" or "transcriptional control region" refers to a polynucleotide encompassing all of the cis-acting sequences necessary for transcription, and may include sequences necessary for regulation. Thus, a transcriptional regulatory region includes at least a promoter sequence, and may also include other regulatory sequences such as enhancers, transcription factor binding sites, polyadenylation signals and splicing signals.

10 "Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter sequence is operably linked to a coding sequence if the promoter sequence promotes transcription of the coding sequence.

15 "Transduction," as used herein, refers to the introduction of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, which methods include, for example, transfection, viral infection, transformation, electroporation and the non-viral gene delivery techniques of the present invention. The introduced polynucleotide may be stably or transiently maintained in the host cell. Stable maintenance typically requires that the introduced polynucleotide either contains
20 an origin of replication compatible with the host cell or integrates into a replicon of the host cell such as an extrachromosomal replicon (e.g. a plasmid) or a nuclear or mitochondrial chromosome.

"Retroviruses" are a class of viruses which use RNA-directed DNA polymerase, or reverse transcriptase, to replicate a viral RNA genome resulting in a double-stranded
25 DNA intermediate which is incorporated into chromosomal DNA of an avian or mammalian host cell. Many such retroviruses are known to those skilled in the art and are described, for example, in Weiss et al., eds, RNA Tumor Viruses, 2d ed., Cold Spring Harbor, New York (1984 and 1985). Plasmids containing retroviral genomes are also widely available, from the American Type Culture Collection (ATCC) and
30 other sources. The nucleic acid sequences of a large number of these viruses are known and are generally available, for example, from databases such as GENBANK.

-11-

A "sequence-specific nucleic acid binding protein" is a protein that binds to nucleic acids in a sequence-specific manner, i.e., a protein that binds to certain nucleic acid sequences (i.e. "cognate recognition sequences", *infra*) with greater affinity than to other nucleic acid sequences. A "sequence-non-specific nucleic acid binding protein" is
5 a protein that binds to nucleic acids in a sequence-non-specific manner, i.e. a protein that binds generally to nucleic acids.

A "cognate" receptor of a given ligand refers to the receptor normally capable of binding such a ligand. A "cognate" recognition sequence is defined as a nucleotide sequence to which a nucleic acid binding domain of a sequence-specific nucleic acid
10 binding protein binds with greater affinity than to other nucleic acid sequences. A "cognate" interaction refers to an intermolecular association based on such types of binding (e.g. an association between a receptor and its cognate ligand, and an association between a sequence-specific nucleic acid binding protein and its cognate nucleic acid sequence).

15 "Gene delivery" is defined as the introduction of targeted nucleic acid into a target cell for gene transfer and may encompass targeting/binding, uptake, transport/localization, replicon integration and expression.

"Lymphocytes" as used herein, are spherical cells with a large round nucleus (which may be indented) and scanty cytoplasm. They are cells that specifically
20 recognize and respond to non-self antigens, and are responsible for development of specific immunity. Included within "lymphocytes" are B-lymphocytes and T-lymphocytes of various classes.

"Lymphohematopoietic stem cells" are cells which are typically obtained from the bone marrow or peripheral blood and which are capable of giving rise, through cell
25 division, to any mature cells of the lymphoid or hematopoietic systems. This term includes committed progenitor cells with significant though limited capacity for self-renewal, as well as the more primitive cells such as those capable of forming spleen colonies in a CFU-S assay, and still more primitive cells possessing long-term and/or multilineage re-populating ability in a transplanted mammalian host.

30 "Lymphohematopoietic cells" include the various mature cells of the lymphoid or hematopoietic systems (including lymphocytes and other blood cells), as well as lymphohematopoietic stem cells.

-12-

A "primary culture of cells" or "primary cells" refer to cells which have been derived directly from in vivo tissue and not extensively passaged. Primary cultures can be distinguished from cell lines and established cultures principally by the retention of a karyotype which is substantially identical to the karyotype found in the tissue from which the culture was derived, and by the cellular responses to manipulations of the environment which are substantially similar to the in vivo cellular responses.

As is described in detail below, the non-viral gene delivery complexes of the present invention comprise gene delivery fusion proteins (GDFPs) that bind targeted nucleic acid through a nucleic acid binding domain (NBD) and facilitate gene delivery through a gene delivery domain (GDD). Each of these domains can comprise a number of different functional components and sub-components. Some of these potential components are summarized in the following list:

NON-VIRAL GENE DELIVERY COMPLEX (the "GDFP/tNA Complex")

1. Gene Delivery Fusion Protein (GDFP)

A. Nucleic Acid Binding Domain (NBD)

- (1) Nucleic Acid Binding (NB) component
- (2) Other possible components (e.g. mediating compression of tNA)

B. Gene Delivery Domain (GDD)

- (1) Binding/Targeting (B/T) component
- (2) Membrane-Disrupting (M-D) component
- (3) Transport/Localization (T/L) component
- (4) Replicon Integration (RI) component

2. Targeted Nucleic Acid (tNA)

- A. Binding sites for the GDFP (see infra)
- B. Sequence of interest (e.g. gene to be delivered)
- C. Other possible sequences (e.g. selectable markers)

Each of these domains and components, as well as additional elements that may be included, are defined and described in detail below.

-13-

The practice of the present invention will employ, unless otherwise indicated, a number of conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature, see, e.g., Kriegler, M. (ed.), "Gene Transfer and Expression, a Laboratory Manual," (1990), W.H. Freeman Publishers; Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual," Second Edition (1989); F.M. Ausubel et al. (eds.), "Current Protocols in Molecular Biology," (1987 and 1993); M.J. Gait (ed.), "Oligonucleotide Synthesis," (1984); R.I. Freshney (ed.), "Animal Cell Culture," (1987); J.M. Miller and M.P. Calos (eds.), "Gene Transfer Vectors for Mammalian Cells," (1987); D.M. Weir and C.C. Blackwell (eds.), "Handbook of Experimental Immunology;" J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach and W. Strober, (eds.), "Current Protocols in Immunology," (1991); and the series entitled "Methods in Enzymology," (Academic Press, Inc.). All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

Illustrations of Type-I Gene Delivery Fusion Proteins

The Gene Delivery Fusion Protein / Targeted nucleic acid Complex (GDFF/tNA)

One concept of the present invention is to create recombinant gene delivery fusion proteins (GDFFs) that are able to bind to a cognate recognition sequence in a targeted nucleic acid (tNA) and facilitate delivery of the tNA into a target cell. The GDFFs bind targeted nucleic acid through a nucleic acid binding domain (NBD) and facilitate gene delivery through a gene delivery domain (GDD).

Thus, in the context of the present invention, targeted nucleic acids can be delivered via one or more steps that are mediated or augmented by GDFFs. In particular, the gene delivery process can include one or more of the following steps: (1) binding and/or targeting of the GDFF/tNA complex to the surface of a target cell; (2) uptake of the tNA (with or without the GDFF) by the target cell; (3) intracellular transport and/or localization of the tNA to an organelle such as a nucleus or mitochondrion; and (4) integration of the tNA into a cellular replicon such as a chromosome. A particular GDFF need not necessarily perform all of these functions.

-14-

For example, a GDFP intended to deliver an expression vector to the nucleus of a cell could be constructed to contain: (i) an NBD capable of binding to a cognate recognition sequence on the expression vector and; (ii) a GDD having only a transport/localization component such as a nuclear localization sequence. Such a GDFP could then be
5 complexed with targeted nucleic acid and introduced into target cells by a transduction method such as electroporation. The GDFP would then facilitate transport/localization to the nucleus, perhaps to a specific site in a replicon, and thus enhance expression of the vector. Alternatively, for example, the aforementioned GDD could be modified to include a binding/targeting component and a membrane-disrupting component. Using
10 such a GDD, the GDFP/tNA complex could be directed to a particular cell type within a population of cells, and uptake of the complex could proceed without the need for, e.g., electroporation. Use of the GDFPs in conjunction with techniques such as electroporation, as in the former example, would of course be more appropriate for in vitro gene delivery. Use of GDFPs as described in the latter example could be readily
15 applied to the delivery of genes either in vitro or in vivo. Similarly, the GDFP/tNA complexes could be used as admixtures with other proteins or simple chemicals that enhance gene delivery. This could include, for example, enhancing the uptake of GDFP/tNA complexes by adding membrane disrupting agents in trans.

Other combinations of components can be prepared (and particular versions of
20 the components can be selected) according to the specific design objectives of the gene delivery scheme. These objectives include, for example, the location of the cells to be targeted, the desired cellular specificity of targeting, and the desired sub-cellular destination of the tNA.

The individual domains and components of the GDFP/tNA complex and their
25 construction and assembly are described in more detail below.

1. The Gene Delivery Fusion Protein (GDFP)

The GDFP comprises two major domains, a nucleic acid binding domain (NBD) and a gene delivery domain (GDD). Each of these major domains comprises one or
30 more components facilitating nucleic acid binding and gene delivery, respectively. These individual components may be derived from naturally-occurring proteins, or they may be synthetic (e.g. an analog of a naturally-occurring component). Typically,

-15-

cloned DNA encoding various components will already be available as plasmids - although it is also possible to synthesize polynucleotides encoding the components based upon published sequence information. Polynucleotides encoding the components can also be readily obtained using polymerase chain reaction (PCR) methodology, as described, for example, by Mullis and Faloona (1987) Meth. Enzymology 155:335.

In the construction of the GDFP, discussed in more detail below, DNA sequences encoding the domains and their various components are preferably fused in-frame so that the GDFP can be conveniently synthesized as a single polypeptide chain (i.e. not requiring further assembly). The various domains and components can also be separated by flexible peptide linker sequences called "flexons" which are defined in more detail above.

A. The Nucleic Acid Binding Domain (NBD)

A nucleic acid binding domain is a length of polypeptide capable of binding (either directly or indirectly) to the targeted nucleic acid (tNA) with an affinity adequate to allow the gene delivery domain of the GDFP to mediate or augment the delivery of the tNA into a target cell. Most conveniently, the NBD will bind directly to the tNA without the need for any intermediary binding element.

In Type-I GDFPs, the NBD contains a sequence-specific binding component that is an analog of a sequence-specific nucleic acid binding protein. In one preferred embodiment of this type, the component allows the nucleic acid binding by the NBD to be sequence-specific with respect to the tNA, in which case the NBD may bind to a specific cognate recognition sequence within the tNA; as is illustrated in Figure 1.

As described herein, one particular advantage of the Type I GDFP approach is that it not only allows the stoichiometric attachment of delivery components to the tNA, but also allows the GDFP to be positioned at pre-determined locations with respect to the tNA. For example, the positioning of NBD cognate recognition sequences in proximity to terminal integrase recognition sequences can facilitate the use of GDFPs to mediate integration, as described below.

The NBD may comprise, for example, a known nucleic acid binding protein, or a nucleic acid binding region thereof. The NBD may also comprise two or more nucleic acid binding regions derived from the same or different nucleic acid binding

-16-

proteins. Such multimerization of nucleic acid binding regions in the NBD can allow for the interaction of the GDFP with the targeted nucleic acid to be of desirable specificity and/or higher affinity. This strategy can be used alone or in combination with multimerization of recognition sequence motifs in the tNA to increase binding avidity, as discussed below.

DNA encoding the NBD domain of the GDFP may be obtained from many different sources. For example, many proteins that are capable of binding nucleic acid have been molecularly cloned and their cognate target recognition sequences have been identified (see, e.g., Mitchell & Tjian, Science 245:371-378, 1989; Pabo & Sauer, Ann. Rev. Biochem. 61:1053-1095, 1992; Harrison, S.C., Nature 353:715-719, 1991; Johnson & McKnight, Ann. Rev. Biochem. 58:799-839, 1989; and references reviewed therein, hereby incorporated by reference). Such sequence-specific binding proteins include, for example, regulatory proteins such as those involved in transcription or nucleic acid replication, and typically have a modular construction, consisting of distinct DNA binding domains and regulatory domains (see, e.g., Struhl, Cell 49:295-297, 1987; Frankel and Kim, Cell 65:717-719, 1991; and Pabo & Sauer, Ann. Rev. Biochem. 61:1053-1095, 1992; and references reviewed therein, hereby incorporated by reference). A number of families of such nucleic acid binding proteins have been characterized on the basis of recurring structural motifs including, for example, Helix-Turn-Helix proteins such as the bacteriophage lambda cI repressor; Homeodomain proteins such as the Drosophila Antennapedia regulator; the POU domain present in proteins such as the mammalian transcription factor Oct2; Zinc finger proteins (e.g. GAL4); steroid receptors; leucine zipper proteins (e.g. GCN4, C/EBP and c-jun); beta-sheet motifs (e.g. the prokaryotic Arc repressor); and other families (including serum response factor, oncogenes such as c-myc, NFkB and rel, and others); see, e.g., Pabo & Sauer, Ann. Rev. Biochem. 61:1053-1095, 1992, and references reviewed therein, hereby incorporated by reference.

For many of these proteins, the nucleic acid binding domains have been mapped in detail; and, for a number of such domains, recombinant fusions with heterologous sequences have been made and shown to retain the binding activities of the parental DNA binding domain. For example, in the case of the yeast-derived transcriptional activator GAL4, the DNA binding domain has been defined, and fusions of this domain

-17-

to heterologous adjoining sequences have been made that retain DNA sequence-specific binding activity (Keegan et al., Science 231:669-704, 1986; Ma & Ptashne, Cell 48:847-853, 1987). This ability to functionally "swap" binding domains has also been shown for a number of other DNA binding proteins, including, for example, the *E. coli* 5 lex A repressor (Brent and Ptashne, Cell 43:729-736, 1985), the yeast transcriptional activator GCN4 (Hope and Struhl, Cell 46:885-894, 1986), the bacteriophage lambda cI repressor (Hu et al., Science 250:1400-1403, 1990), the mammalian transcription factors Sp1 (Kadonaga et al., Cell 51:1079-1090, 1987) and C/EBP (Agre et al., Science 246:922-926, 1989). Similarly, functional swapping has been reported in the 10 nuclear DNA-binding steroid hormone receptors (see, e.g., Green and Chambon, Nature 325:75-78, 1987). See also, e.g., Klug & Rhodes, Trends Biochem. Sci. 12:464-471, 1987; Berg, Cell 57:1065-1068, 1989; Waslylyk et al., Eur. J. Biochem. 211:7-18, 1993; Faisst & Meyer, Nucl. Acids Res. 20:3-26, 1992; Struhl, Trends Biochem. Sci. 14:137-140, 1989; and Nelson & Sauer, Cell 42:549-558, 1985.

15 Sequence-specific nucleic acid binding proteins can exhibit a range of binding affinities to different cognate nucleic acid sequences in vitro (see, e.g., Vashee et al., J. Biol.Chem 268:24699-24706, 1993).

Virally encoded nucleic acid binding proteins can also be used in the present invention. These include, for example, the adenovirus E2A gene product, which can 20 bind single-stranded DNA, double-stranded DNA and also RNA (Cleghon et al., Virology 197:564-575, 1993, and references cited therein); the retroviral IN proteins (Krogstad & Champoux, J. Virol 64: 2796, 1990); the AAV rep 68 and 78 proteins (Owens et al., J. Virol 67: 997, 1993); and the SV40 T antigen (Arthur et al., J. Virol., 62:1999-2006, 1988). The cellular p53 gene product, which binds T antigen, is 25 also a DNA binding protein (Funk et al., Mol. Cell Biol., 12:2866-2871, 1992).

Similarly, RNA binding proteins have been identified and their inclusion in the NBD would associate the GDFP with a targeted RNA and thereby achieve RNA delivery mediated by the gene delivery domain of the GDFP. RNA binding proteins that can be used in the context of the present invention include, for example, the Tat 30 and Rev proteins of HIV; see, e.g., Tiley et al., P.N.A.S. 89:758-762, 1992; and Cullen et al., Cell 73:417-420, 1993. Similarly, cellular RNA binding proteins, such

-18-

as the interferon-inducible 9-27 gene product (Constantoulakis et al., Science 259:1314-1318, 1993), can also be used.

Nucleic acid binding domains of Type-I GDFPs can also contain (in addition to a component derived from a sequence-specific nucleic acid binding protein) one or more components that are derived from sequence-non-specific nucleic acid binding proteins. Such sequence-non-specific binding proteins include, for example, histones (von Holt, Bioassays 3:120-124, 1986; Rhodes, Nucleic Acids Res. 6:1805-1816, 1979; Rodriguez et al., Biophys. Chem. 39:145-152, 1991); proteins such as nucleolin (Erard et al., Eur. J. Biochem. 191:19-26, 1990); polybasic polypeptide sequences such as poly-L-lysine (Li et al., Biochemistry, 12:1763-1772 1973; Weiskopf and Li, Biopolymers 16:669-684, 1977), avidin (Pardridge & Boado, F.E.B.S. Lett. 288:30-32, 1991); the non-histone high mobility group proteins and other proteins (see, e.g., Pabo & Sauer, Ann. Rev. Biochem. 61:1053-1095, 1992, and references reviewed therein); that interact non-specifically with nucleic acids. Other proteins binding nucleic acid in a sequence-non-specific fashion include retroviral nucleocapsid (NC) proteins (see, e.g., Gelfand et al., J. Biol. Chem., 268:18450-18456, 1993).

B. The Gene Delivery Domain (GDD)

The GDD portion of the GDFP contains one or more polypeptide regions that mediate or augment the efficiency of gene delivery. Such sequences may include, for example, binding/targeting components, membrane-disrupting components, transport/localization components, and replicon integration components, as discussed below.

A particular GDD need not contain a component representing each of the aforementioned types. Conversely, a GDD may contain more than a single component of a given type to obtain the desired activity. Moreover, a particular segment of a GDD might serve the function of two or more of these components. For example, a single region of a polypeptide might function both in binding to a cell surface and in disruption of the membrane at that surface.

30

-19-

(1) Binding/Targeting (B/T) Components

Binding/targeting components are regions of polypeptides that mediate binding to cellular surfaces (which binding may be specific or non-specific, direct or indirect). Any protein that can bind to the surface of the desired target cell can be employed as a source of B/T components. Such proteins include, for example, ligands such as cytokines that bind to particular cell surface receptors, antibodies, lectins, viral binding proteins, cellular adhesion molecules, and any other proteins that associate with cellular surfaces. The "receptors" for these binding proteins include but are not limited to proteins. Moreover, the receptors may, but need not, be specific and/or restricted to certain cell types. Essentially, the B/T components can be prepared from any ligand that binds to a cell surface molecule.

By way of illustration, one group of proteins from which the B/T components can be derived are cytokines. Cytokines are intercellular signalling molecules, the best known of which are involved in the regulation of mammalian somatic cells. Several families of cytokines, both growth promoting and growth inhibitory in their effects, have been characterized. Thus, a B/T component can comprise an amino acid sequence containing at least that portion of a cytokine polypeptide that is required for binding to receptors for the cytokine on the surface of mammalian cells, or a mutein of such a portion of a cytokine polypeptide. A B/T component derived from a cytokine can, but need not, also contain the portion of the cytokine that is involved in "cytokine effector activity," as described below.

Examples of cytokines that can be used in the present invention include, for example, interleukins (such as IL-1 α , IL-1 β , IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9 (P40), IL-10, IL-11, IL-12 and IL-13); CSF-type cytokines such as GM-CSF, G-CSF, M-CSF, LIF, EPO, TNF- α and TNF- β); interferons (such as IFN- α , IFN- β , IFN- γ); cytokines of the TGF- β family (such as TGF- β 1, TGF- β 2, TGF- β 3, inhibin A, inhibin B, activin A, activin B); chemotactic factors (such as NAP-1, MCP-1, MIP-1 α , MIP-1 β , MIP-2, SIS β , SIS δ , SIS ϵ , PF-4, PBP, γ IP-10, MGSA); growth factors (such as EGF, TGF- α , aFGF, bFGF, KGF, PDGF-A, PDGF-B, PD-ECGF, INS, IGF-I, IGF-II, NGF- β); α -type intercrine cytokines (such as IL-8, GRO/MGSA, PF-4, PBP/CTAP/ β TG, IP-10, MIP-2, KC, 9E3); and β -type intercrine cytokines (such as MCAF, ACT-2/PAT 744/G26, LD-78/PAT 464, RANTES, G26, I309, JE, TCA3,

-20-

MIP-1 α , β , CRG-2). A number of other cytokines are also known to those of skill in the art. The sources, characteristics, targets and effector activities of these cytokines have been described and, for many of the cytokines, the DNA sequences encoding the molecules are also known; see, e.g., Van Snick, J. et al. (1989) *J. Exp. Med.* 169: 363-368; Paul, S.R. et al. (1990) *Proc. Natl. Acad. Sci. USA* 87: 7512-7516; Gately, M.K. et al. (1991) *J. Immunol.* 147: 874-882; Minty, A., et al. (1993) *Nature* 362: 248; and the reviews by Arai, K., et al. (1990) *Annu. Rev. Biochem.* 59:783-836; and Oppenheim, J.J., et al. (1991) *Annu. Rev. Immunol.* 9:617-48; Waldman, T.A. (1989) *Annu. Rev. Biochem.* 58:875-911; Beutler, B., et al. (1988) *Annu. Rev. Biochem.* 57:505-18; Taniguchi, T. (1988) *Annu. Rev. Immunol.* 6:439-64; Paul, W.E. et al., (1987) *Annu. Rev. Immunol.* 5:429-59; Pestka, S. et al., (1987) *Annu. Rev. Biochem.* 56:727-77; Nicola, N.A. et al. (1989) *Annu. Rev. Biochem.* 58:45-77; and Schrader, J.W. (1986) *Annu. Rev. Immunol.* 4:205-30; and the particular publications reviewed and/or cited therein, which are hereby incorporated by reference in their entirety.

Many of the DNA sequences encoding cytokines are also generally available from sequence databases such as GENBANK. Typically, cloned DNA encoding such cytokines will already be available as plasmids - although it is also possible to synthesize polynucleotides encoding the cytokines based upon the published sequence information. Polynucleotides encoding the cytokines can also be obtained using polymerase chain reaction (PCR) methodology, as described, for example, by Mullis and Faloona (1987) *Meth. Enzymology* 155:335. The detection, purification, and characterization of cytokines, including assays for identifying new cytokines effective upon a given target cell type, have also been described in a number of publications, including, e.g., Clemens, M.J. et al. (eds.) (1987) "Lymphokines and Interferons," IRL Press, Oxford; and DeMaeyer, E., et al. (1988) "Interferons and Other Regulatory Cytokines," John Wiley & Sons, New York; as well as the references referred to above.

The ligands suitable for targeting a particular sub-population of cells will be those which bind to receptors present on cells of that sub-population. Again, taking cytokines as an example, the target cells for a large number of these molecules are already known, as noted above; and, in many cases, the particular cell surface receptors for the cytokine have already been identified and characterized; see, e.g., the

-21-

publications referred to above. Typically, the cell surface receptors for cytokines are transmembrane glycoproteins that consist of either a single chain polypeptide or multiple protein subunits. The receptors generally bind to their cognate ligands with high affinity and specificity, and may be widely distributed on a variety of somatic cells, or quite specific to given cell subsets. The presence of cytokine receptors on a given cell type can also be predicted from the ability of a cytokine to modulate the growth or other characteristics of the given cell; and can be determined, for example, by monitoring the binding of a labeled cytokine to such cells; and other techniques, as described in the references cited above.

Thus, for example, a large number of cytokine receptors have been characterized and many of these are known to belong to receptor families which share similar structural motifs; see, e.g., the review by Miyajima, A., et al., *Ann. Rev. Immunol.* 10:295-331 (1992), and the publications reviewed therein, hereby incorporated by reference. Type-I cytokine receptors (or hematopoietic growth factor receptors) include, for example, the receptors for IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, GM-CSF, G-CSF, EPO, CNTF and LIF. Type-II cytokine receptors include, for example, the receptors for IFN- α , IFN- β and IFN- γ . Type-III cytokine receptors include, for example, the receptors for TNF- α , TNF- β , FAS, CD40 and NGF. Type-IV cytokine receptors (immunoglobulin-like, or "Ig-like," receptors) include the receptors for IL-1; and the receptors for IL-6 and G-CSF (which have Ig-like motifs in addition to the Type-I motif). These receptor families are described for example, in Smith et al., *Science* 248: 1019-1023, 1990; Larsen et al., *J. Exp. Med.*, 172: 1559-1570, 1990; McMahan et al., *EMBO J.* 10:2821-2832, 1991; and in the reviews by Cosman et al., *Trends Biochem Sci* 15: 265-269, 1990; and Miyajima, A., et al., *Ann. Rev. Immunol.* 10:295-331 (1992), and the publications reviewed therein, all of which are hereby incorporated by reference. As new cytokines are characterized, these can be employed in the present invention as long as they exhibit the desired binding characteristics and specificity. The identification and characterization of cytokines, and the use of assays to test the ability of cytokines to activate particular target cells, are known in the art; see, e.g., Clemens, M.J. et al. (eds.) (1987) "Lymphokines and Interferons," IRL Press, Oxford; and DeMaeyer, E., et al. (1988) "Interferons and

-22-

Other Regulatory Cytokines," John Wiley & Sons, New York; as well as the references referred to above.

The choice of a particular ligand will depend on the presence of cognate receptors on the desired target cells. It may also depend on the corresponding absence
5 of cognate receptors on other cells which it may be preferable to avoid targeting. With the cytokines, for example, the role of particular molecules in the regulation of various cellular systems is well known in the art. In the hematopoietic system, for example, the hematopoietic colony-stimulating factors and interleukins regulate the production and function of mature blood-forming cells. Lymphocytes are dependent upon a
10 number of cytokines for proliferation. For example, cytotoxic T lymphocytes (CTLs) are dependent on helper T (T_H) cell-derived cytokines, such as IL-2, for growth and proliferation in response to foreign antigens. (Zinkernagel and Doherty, Adv. Immunol. 27:51, 1979; Male et al., Advanced Immunology, Chap. 7, Gower Publ., London, 1987; Jacobson et al., J. Immunol. 133:754, 1984). IL-2, for example, is a
15 potent mitogen for CTLs (Gillis and Smith, Nature 268:154, 1977), and the combination of antigen and IL-2 cause proliferation of primary $CD4^+$ T cells in vitro. The importance of IL-2 for the growth and maintenance of the $CD8^+$ CTL in vivo has been documented in models of adoptive immunotherapy in which the therapeutic efficacy of transferred anti-retroviral $CD8^+$ cells is enhanced on subsequent
20 administration of IL-2 (Cheever et al., J. Exp. Med. 155:968, 1982; Reddehase et al., J. Virol. 61:3102, 1987). IL-4 and IL-7 are also capable of stimulating the proliferation of mature $CD8^+$ CTL (Alderson et al., J. Exp. Med. 172:577, 1990). In the case of IL-2, the IL-2 receptors are expressed on T-cells, B-cells, natural killer cells, glioma cells and cells of the monocyte lineage (Smith, Science 240:1169, 1988).
25 However, the greatest level of high affinity IL-2 receptor expression is observed in activated T-cells (Waldemann, Ann. Rev. Biochem. 58:875, 1989). The IL-2 receptor complex consists of three protein components, a low affinity receptor, α , Tac or p55 (Leonard et al., Nature 311:626, 1984), an intermediate affinity receptor, β or p70 (Hatakeyama, Science 244:551, 1989), and a signal transduction protein, γ or p64,
30 which interacts with the p70 receptor subunit (Takeshita et al., Science 257:379, 1992). The combination of the α and β subunits together make up a high affinity form of the

-23-

IL-2 receptor (Hatakeyama, Science 244:551, 1989); α - β - γ combinations appear to have the highest affinity (Asa et al., P.N.A.S. 90:4127-4131, 1993).

Thus, for example, a GDFP including IL-2 can be used to target gene delivery specifically to activated T lymphocytes which express high levels of α - β - γ high affinity
5 receptors. The cellular targets of a large number of the other cytokines are known and described in the reviews and other references cited above. Furthermore, following the approaches described in those references, any particular cell population or sub-population can be readily assayed for sensitivity to a given cytokine.

The choice of a particular ligand may also be influenced by other activities that
10 may be possessed by the ligand (besides binding to the cell surface). For example, GDFPs having B/T components derived from cytokines may possess cytokine effector activity that can be used to modulate the targeted cells in accordance with the activity of the cytokine. Typically, GDFPs of this type will be prepared by incorporating the entire cytokine coding sequence into a polynucleotide encoding the GDFP; although it
15 will also be possible to remove portions of the cytokine sequence which are neither required for binding to the receptor nor essential for cytokine effector activity. In such cases, the GDFPs can provide a combination of activities comprising: (i) binding to specific target cells; (ii) delivery of targeted nucleic acid into the targeted cells; and (iii) cytokine modulation of the cells thus targeted. Such a combination of activities
20 will allow, for example, the transduction of particular cells to be coupled to the proliferation of the transduced cells. This will be generally advantageous in the context of gene delivery since it can be used to promote the proliferation of the targeted cells in a given cell population; and will be particularly advantageous for in vivo gene delivery where it may be otherwise problematic or impossible to induce the targeted
25 cells to divide, which may be necessary for efficient stable incorporation of the transferred gene.

In some cases, it will be preferable to make use of the receptor binding potential of a ligand such as a cytokine without concomitant effector activity. This may be the case, for example, when a cytokine with suitable receptor binding properties has a
30 negative or unwanted effect on target cell activity. GDFPs of this type can be prepared, for example, from cytokine sequences in which the domain responsible for effector activity has been mutationally altered by, e.g., substitution, insertion or

-24-

deletion. For example, IL-2 has been subjected to deletion analysis to identify which portions of the sequence are involved in receptor binding and which are critical for cytokine effector activity; see, e.g., Brandhuber, B.J. et al., J. Biol. Chem. 262:12306-308, 1987; Brandhuber, B.J. et al., Science 238:1707-09, 1987; Zurawski, S.M. et al., EMBO J. 7:1061-69, 1988; and Arai, K., et al., Annu. Rev. Biochem. 59:783-836, 1990. The receptor binding and effector domains of a number of other cytokines have similarly been characterized; see Arai et al., id, and other reviews and references cited therein.

The rapidity with which novel ligands and their cognate receptors have recently been molecularly cloned has generated a wide array of these molecules. In particular, the combination of direct cDNA expression cloning and screening assays for either induction of proliferation or binding to specific cell surface receptors on target cells has led to many new molecules being cloned (see, for example, Cosman et al., Trends Biochem Sci 15: 265-269, 1990). The advent of these technologies will undoubtedly lead to the cloning of more ligands, including cytokines and other proteins that bind to cells, which, on the basis of their binding characteristics and specificity may be used in the context of the present invention as the B/T component of the GDFP. B/T components derived from the flk-2/flt-3 ligand (Lyman et al., Cell 75:1157-1167, 1993) will be of interest because the cytokine binds specifically to a receptor, flk-2/flt-3, which is expressed on early hematopoietic cells (Matthews, W. et al., Cell 65:1143, 1991; and Small et al., P.N.A.S. 91:459-463, 1994). In the context of the present invention, GDFPs comprising a B/T component derived from the flk-2 ligand could thus be used to direct gene delivery to lymphohematopoietic stem cells.

While the foregoing principles have been illustrated using cytokines as a convenient example, these principles are also applicable to other ligands capable of binding to cell surfaces, including for example, antibodies, lectins, viral binding proteins, cellular adhesion molecules, and any other proteins that associate with cellular surfaces.

For example, a large number of antibodies to cell surface antigens have been identified and described. Antibodies to leukocytes have been well characterized and classified as the "CD" series of antigens; see, e.g., Coligan, J. et al. (ed.), "Current Protocols in Immunology," Current Protocols, 1992, 1994. Moreover, techniques for

-25-

the isolation of new antibodies specific for a particular target cell are routine in the art. Useful antibodies will be those which interact with antigens on the surface of the desired target cells. Antibody/antigen binding can be readily determined and monitored by flow cytometry or other immunochemical detection methods.

5 Of particular interest are antigens that are exclusively or preferentially expressed on the surface of particular target cells. For example, the CD34 antigen is expressed on human lymphohematopoietic stem cells (Andrews et al., Blood 80:1693-1701, 1992).

 Transferrin, (see, e.g., Zenke, M. et al., P.N.A.S. 87:3655-3659, 1990), can
10 also be used as a B/T component in the context of the present invention.

 Targeting to certain cells, for example respiratory epithelial cells, can also take place via immunoglobulin (Ig) receptors (see, e.g., Ferkol, T., J. Clin. Invest. 92:2394-2400 (1993)).

 The GDFPs of the present invention can also be chemically modified, for
15 example by the addition of lactose to target the GDFP to asialoglycoprotein receptors and thus to hepatocytes of the liver (see, e.g., Neda, H. et al., J. Biochem. 266:14143-14146, 1991).

 Another group of proteins from which the B/T components can be derived are lectins. A number of such molecules, and their cognate receptors, have been identified
20 and characterized (see, e.g., the review by Lis & Sharon, Ann. Rev. Biochem. 55:35-67, 1986; and publications cited therein).

 Proteins capable of targeting the GDD and thus the GDFP/tNA complex to cell surfaces can also be derived from viruses. Many such viral proteins capable of binding to cells have been identified, including, for example, the well-known envelope ("env")
25 proteins of retroviruses; hemagglutinin proteins of RNA viruses such as the influenza virus; spike proteins of viruses such as the Semliki Forest virus (Kielian and Jungerwirth (1990) Mol. Biol. Med. 7:17-31); and proteins from non-enveloped viruses such as adenoviruses (see, e.g., Wickham et al., Cell 73:309-319, 1993).

 As an illustrative example, in the murine leukemia virus (MuLV) system, it is
30 well known that the amino-terminal region of the gp70 molecule is involved in binding to cell surface receptors, see, e.g., Heard and Danos, J. Virol. 65: 4026-4032, 1991. Battini et al., J. Virol. 66: 1468-1475 (1992) have also reported that portions of the

-26-

amino-terminal region of gp70 can be exchanged in order to switch binding to different MuLV env receptors without interfering with the ability of the protein to interact with p15E TM protein (and, thereby, to mediate viral uptake); see also Weiss, R. et al. in Weiss, R. et al. (eds.), RNA Tumor Viruses, Cold Spring Harbor, New York (1984 and 1985). Similarly, in the human immunodeficiency virus (HIV) system, mutational analysis of gp120 has identified portions of the molecule which are critical for binding to the CD4 receptor, see, e.g., Kowalski, M. et al., Science 237:1351-1355, 1987. Yet another approach to identify a region critical for receptor binding is as follows: an antibody known to inhibit binding can be used to immuno-affinity purify a cleavage fragment of the viral binding protein; which fragment is then partially sequenced to identify the corresponding domain of the viral binding protein, see, e.g., Laskey, L.A. et al., Cell 50:975-985, 1987. Such techniques can be employed in the present invention to generate GDFPs in which the M-D component remains capable of mediating uptake of the GDFP/tNA complex (as described below), but the specificity of binding is principally determined by the presence of, e.g., cognate cytokine receptors corresponding to a portion of the B/T component, rather than viral binding protein receptors.

Another illustrative example of a viral protein that can be used is the G protein of VSV, which has been utilized to target infection by retroviral vectors; see, e.g., Emi et al., J. Virol., 65:1202-1207, 1991.

Another group of proteins from which the B/T components can be derived are cellular adhesion molecules. A number of such molecules, and their cognate receptors, have been identified and characterized (see, e.g., Springer, T., Nature 346:425-434, 1990, and publications cited therein).

25

(2) Membrane-Disrupting (M-D) Components

Membrane-disrupting components are protein sequences capable of locally disrupting cellular membranes such that the GDFP/tNA complex can traverse a cellular membrane.

30 M-D components facilitating uptake of the GDFP-targeted nucleic acid complex by target cells are typically membrane-active regions of protein structure having a

-27-

hydrophobic character. Such regions are typical in membrane-active proteins involved in facilitating cellular entry of proteins or particles.

For example, viruses commonly enter cells by endocytosis and have evolved mechanisms for disrupting endosomal membranes. Many enveloped viruses encode surface proteins capable of disrupting cellular membranes including, for example, retroviruses, influenza virus, Sindbis virus, Semliki Forest virus, Vesicular Stomatitis Virus, Sendai virus, Vaccinia virus, and mouse hepatitis virus; see e.g., Kielian and Jungerwirth, *Mol. Biol. Med.*, 7:17-31, 1990; and Marsh & Helenius, *Adv. Virus Res.*, 36:107-151, 1989. The mechanism for viral entry, in which a viral binding protein binds to a specific cell surface receptor and subsequently mediates virus entry, frequently by means of a hydrophobic membrane-disruptive domain, is a common theme among enveloped viruses, including influenza virus, and many such molecules are known to those skilled in the art, see, e.g., Hunter and Swannstrom, *Curr. Top. Micro. and Immunol.* 157:187, 1990; and the review by White, J., *Science* 258:917-924, 1992; and publications reviewed therein.

By way of illustration, the M-D components of the present invention can thus be derived from a portion of a viral binding protein that is normally involved in mediating uptake of the virus into a host cell, or a mutein of such a portion of a binding protein. The portion of the GDFP that may be derived from such a viral binding protein may, but need not, also contain the portion of the binding protein that causes the viral particle to associate with a specific receptor on a target cell (which latter portion may thus function as a B/T component, as described above). A large number of viruses have been characterized and, for many of these, the nucleotide sequence of the viral genome has been published. The binding proteins encoded by various viruses generally share functional homology, even though there may be considerable variation among the primary amino acid sequences. Using the retroviruses to illustrate, the native env gene product is typically a polyprotein precursor that is proteolytically cleaved during transport to the cell surface to yield two polypeptides: a glycosylated polypeptide on the external surface (the "SU" protein) and a membrane-spanning or transmembrane protein (the "TM" protein); see, e.g., Hunter, E. and R. Swannstrom, *Curr. Topics Microbiol. Immunol.* 157:187-253, 1990. The SU proteins are responsible for binding to specific receptors on the surface of target cells as a first step in the infection

-28-

process. The TM proteins, as well as associating with viral core proteins through their C-terminal ends, are responsible for a critical membrane fusion event which takes place after binding and allows entry of the virus into the cell (See, e.g., Hunter and Swanstrom (1990) *Curr. Top. Micro. and Immunol.* 157:187; Kielian and Jungerwirth (1990) *Mol. Biol. Med.* 7:17-31; and Marsh & Helenius (1989) *Adv. Virus Res.*, 36:107-151). The membrane fusion event is accomplished by a hydrophobic polypeptide sequence present at the amino terminus of the TM protein. Examples of these pairs of SU and TM proteins and the viruses which produce them are: gp52 and gp36 from mouse mammary tumor virus (Racevskis, J. et al., *J. Virol.* 35:937-48, 1980); gp85 and gp37 from Rous sarcoma virus (Hunter et al., *J. Virol.* 46:920, 1983); gp70 and p15E from Moloney murine leukemia virus (Koch et al., 49:828, 1984); gp70 and gp20 from Mason Pfizer monkey virus (Bradac, J. et al., *Virology* 150:491-502, 1986); gp120 and gp41 from human immunodeficiency virus (Kowalski, M. et al., *Science* 237:1351-1355, 1987); and gp46 and gp21 from human T-Cell leukemia virus (Seiki et al., *Proc. Natl. Acad. Sci.* 80:3618, 1983); and others described in the references cited herein. The functional similarity among these types of proteins, is further illustrated by the well-documented phenomenon of "pseudotyping," in which the core proteins and nucleic acid are provided by a first virus and the envelope proteins (determining host range) are provided by a different virus (see, e.g., Vile et al., *Virology* 180:420, 1991; Miller et al, *J. Virol.* 65:2220, 1991; and Landau et al., *J. Virol.* 65:162, 1991). Examples of retroviruses which can be used to derive fragments for use in the present invention include murine retroviruses such as Harvey murine sarcoma virus (Ha-MSV), Kirsten murine sarcoma virus (Ki-MSV), Moloney murine sarcoma virus (Mo-MSV), various murine leukemia viruses (MuLV), mouse mammary tumor virus (MMTV), murine sarcoma virus (MSV) and rat sarcoma virus (RaSV); bovine leukemia virus (BLV); feline retroviruses such as feline leukemia virus (FeLV) and feline sarcoma virus (FeSV); primate retroviruses such as baboon endogenous virus (BaEV), human immunodeficiency viruses (HIV-I and HIV-II), human T-cell leukemia viruses (HTLV-I and HTLV-II), Gibbon ape leukemia virus, Mason Pfizer monkey virus (M-PMV), simian immunodeficiency virus (SIV) and simian sarcoma virus (SSV); various lentiviruses; and avian retroviruses such as avian erythroblastosis virus, avian leukosis virus (ALV), avian myeloblastosis virus, avian sarcoma virus (ASV),

avian reticuloendotheliosis-associated virus (REV-A), Fujinami sarcoma virus (FuSV), spleen necrosis virus (SNV) and Rous sarcoma virus (RSV). Many other suitable retroviruses are known to those skilled in the art and a taxonomy of retroviruses is provided by Teich, pp. 1-16 in Weiss et al., eds, RNA Tumor Viruses, 2d ed., Vol.2, Cold Spring Harbor, New York. Plasmids containing retroviral genomes are also
5 widely available from the ATCC and other sources.

Infectious virions have also been produced when non-retroviral binding proteins, such as the G protein of vesicular stomatitis virus or the hemagglutinin of influenza virus, have been pseudo-typed onto retrovirus cores (see, e.g., Emi et al., J. Virol. 65:1207, 1991; and Dong et al., J. Virol. 66:7374, 1992). These latter examples
10 indicate that there are functional commonalities between various viruses and their mode of entry into cells which will allow the use of viral binding proteins from a variety of sources. Influenza hemagglutinin has also been reported to enhance the uptake of poly-L-lysine-based chemical conjugates (Wagner et al., P.N.A.S. 89:7934-7938, 1992).

The sequences of a large number of viral binding proteins are known, and are generally available from sequence databases such as GENBANK. Furthermore, polynucleotides encoding viral binding proteins can be readily obtained from viral particles themselves. Also, since many different genes encoding viral binding proteins have been cloned and characterized, plasmids containing DNA encoding the binding
15 proteins are available from a number of different sources. Polynucleotides encoding viral binding proteins can also be obtained using polymerase chain reaction (PCR) methodology, as described, for example, by Mullis and Faloona (1987) Meth. Enzymology 155:335.

As an illustrative embodiment of the present invention, a GDFP may comprise a region of a gene encoding a viral binding protein including a B/T component, in which
25 case the GDFP can be used to target cells including those normally susceptible to the virus from which the gene was derived. In other embodiments of the present invention, the targeting may be restricted to cells bearing receptors for other types of ligands, discussed above under the description of the B/T component. For example, where an M-D component is derived from a viral binding protein that retains the ability
30 to bind to the viral receptor, but it is desirable to limit targeting to cells bearing, e.g., an appropriate cytokine receptor, there are several approaches that can be used to

-30-

achieve such specificity. One approach is to utilize a B/T component which is based on a cytokine with a very high binding affinity for the desired target cells compared to the binding affinity of a domain derived from a viral binding protein for the native viral binding protein receptors. Since many of the cytokines are known to exhibit very high affinity binding to their receptors, and since it will be feasible, for example, to base the M-D component on a lower-binding-affinity viral binding protein, targeting can be effectively focused upon those cells bearing a cognate cytokine receptor. Another suitable approach to limiting binding is to derive the M-D component from a mutant viral binding protein in which the mutation has disrupted the ability of the protein to engage in binding via the native viral binding protein receptor but has not interfered with the ability of the viral binding protein to mediate viral uptake. Plasmids encoding such mutant viral binding proteins are available in the art; and it will also be well within the ability of one skilled in the art to prepare new versions of such viral binding protein mutants by deleting portions of the coding region or by introducing amino acid substitutions into the coding sequence as described above.

While the foregoing principles have been illustrated using viral proteins as a convenient example, these principles are also applicable to other polypeptides capable of disrupting cellular membranes (see, e.g., the review by White, J., Science 258:917-924, 1992, and publications reviewed therein).

Other domains that are functionally and/or structurally analogous can be derived from various viral, prokaryotic or eukaryotic sources. As a further specific example, bacterial toxins such as diphtheria toxin have a specific domain with a highly alpha-helical structure and a hydrophobic character (known as the "TM" domain in the case of diphtheria toxin) that becomes protonated at low pH and disrupts cellular membranes, facilitating entry of the toxin into the cells (see, e.g., Choe et al. (1992) Nature 357:216-222; vanderSpeck et al, J. Biol Chem 268: 12077-12082, 1993; and Parker & Pattus, Trends Biochem. Sci. 18:391-395, 1993). Toxins such as *Pseudomonas* exotoxin A have a similar membrane-disrupting domain (see, e.g., Strom et al., Ann. N.Y. Acad. Sci. 636:233-250, 1991). Similar M-D components can be derived from other bacterial toxins such as hemolysin (Suttorp et al., J. Exp. Med., 178:337-341, 1993). As described herein, inclusion of such membrane disruptive

-31-

components in the GDFP would facilitate membrane disruption and entry of the GDFP-tNA complex into the target cells.

Cytolytic pore-forming proteins, such as streptolysin O, perforins expressed by cytotoxic T lymphocytes, and *S.aureus* alpha toxin also have the ability to disrupt membranes (see, e.g., Ojcius and Young, Trends Biochem. Sci., 16:225-230, 1991; 5 Suttorp et al., J. Exp. Med., 178:337-341, 1993). Streptolysin O has been shown to facilitate uptake of DNA by cultured cells when added to the culture medium (Barry et al., Biotechniques 15:1018-1020). There are many bacterial cytolysins which have the capability to induce membrane disruption (see, e.g., Braun and Focareta., Crit. Rev. 10 Microbiol. 18:115-158, 1991; and van der Goot et al., Nature 354:408-411, 1991). Membrane disruption often occurs by means of a pH-induced hydrophobic change in the protein, but this can also occur by enzymatic means, such as those involving phospholipases (see, e.g., Braun and Focareta., Crit. Rev. Microbiol. 18:115-158, 1991 (and references cited therein); and London, Mol. Microbiol. 6:3277-3282, 1992). 15 Where a pH shift is required to induce the membrane disruption function, there are several ways in which this can be achieved. For example, the GDFP/tNA complex may be taken up through acid endosomes; or the pH of the extracellular medium may be transiently lowered to mediate activation of the membrane disruption function. In some cases (diphtheria toxin for example), enzymatic nicking of the membrane active 20 component prior to an induced pH change in the surrounding medium is believed to promote membrane disruption (see, e.g., Sandvig and Olsnes, J. Cell Biol. 87:828-832, 1980; Moskaug et al., J. Biol. Chem. 263:2518-2525, 1988; and Zalman and Wisnieski, Proc. Natl. Acad. Sci. 81:3341-3345, 1984). Well-known enzymes such as trypsin and urokinase have been successfully used to provide the nicking activity in 25 vitro (see, e.g., Williams, D.P., et al., J. Biol. Chem. 265:20673-20677, 1990). Enzymes capable of providing the nicking activity are also known to be found on cellular surfaces (see, e.g., Williams, D.P., et al., id.). Exemplary construction and characterization of GDFPs containing the diphtheria toxin transmembrane region are described below in Example 8.

30 Other sources of M-D components include bacterial proteins that promote entry of organisms into cells, such as the 52kD entry protein of *Mycobacterium tuberculosis* (Arruda et al., Science 261:1454-1457, 1993); the internalin protein of *Listeria*

monocytogenes (Portnoy et al., Inf. Imm. (U.S.) 60:1263-1267); and the invasins protein of *Yersinia enterocolitica* (Young et al., J. Cell Bio., 116, 197-207), among others.

Synthetic analogs of membrane-disrupting domains can also be made. See, e.g.,
5 Kaiser and Kezdy, Science 223:249-255, 1984.

(3) Transport/Localization (T/L) Components

Transport/localization components mediate or augment the transport and/or localization of the GDFP/tNA complex to a particular sub-cellular compartment such as
10 the nucleus or mitochondrion.

A number of sequences that mediate transport and/or localization of proteins have been identified. These include, by way of illustration, the nuclear localization sequence (nls) of SV40 T antigen (Colledge, et al., Mol. Cell Bio. 6:4136-4139, 1986); and the HIV matrix protein (Bukrinsky et al., Nature 365:666-669, 1993). These are
15 typically short basic peptide sequences, and may also be bipartite basic sequences (see, e.g., Garcia-Bustos et al., Biochim. Biophys. Acta 1071:83-101, 1991; and Robbins et al., Cell 64:615-623, 1991). Nuclear localization sequences have been fused to heterologous proteins and shown to confer on them the property of nuclear localization (see, e.g., Biocca et al., EMBO J. 9:101-108, 1990). In the case of the human
20 estrogen receptors, for example, fusion proteins traffic to the nucleus in an estrogen-dependent fashion (Ishibashi et al., J. Biol. Chem. 269:7645-7650, 1994). These sequences can be readily incorporated into the GDD by recombinant DNA methodology to facilitate nuclear localization of the desired GDFP/tNA complex. GAL4 has also been shown to possess nuclear localization properties in yeast (see, e.g., Silver, et al.,
25 P.N.A.S. 81:5951-5955, 1984), and thus, as a component of a GDFP, GAL4 may be used as both an NBD and a GDD with a role in transport/localization.

(4) Replicon Integration (RI) Components

Replicon integration components mediate or augment integration of the targeted
30 nucleic acid into a replicon of the target cell, such as a chromosome. In many instances in gene transfer and gene therapy it is advantageous to obtain stable integration of transferred DNA into the genome of the target cell. The GDFP can

-33-

facilitate such integration. Also, as described herein, a particular advantage of the Type I GDFP approach is that it not only allows the stoichiometric attachment of delivery components to the tNA, but also allows the GDFP to be positioned at pre-determined locations with respect to the tNA. In the case of a replicon integration component such as an integrase, the GDFP can be positioned in proximity to terminal integrase recognition sequences as a means of facilitating integration, as described in more detail below.

DNA-protein interactions can mediate integration of DNA into the mammalian genome. For example, the integration of all known retroviruses takes place in an enzymatic reaction that makes an endonucleolytic cleavage of the host DNA and ligates the reverse-transcribed retroviral genome to the free ends of the host cell DNA. This reaction is mediated by the retroviral integrase (or "IN") protein, and it is well known that the IN protein interacts with a minimal number of bases present on the ends of the pre-integrative viral genome to achieve integration. Indeed, DNA sequences bearing the IN sequence recognition motif can be inserted into free DNA in vitro by purified IN proteins (see, e.g., Bushman et al., Science 249:1555-1558, 1990; and Katz et al., Cell 63:87-95, 1990; see also, Brown et al. Cell 49:347-356, 1987; and Roth et al., Cell 58:47-54, 1989). For example, the MLV, HIV and RSV IN proteins are each known to interact with a distinct short IN sequence recognition motif present at each end of the linear pre-integrative viral DNA substrate to mediate its integration into the host cell replicon. In vitro integration mediated by purified IN protein has been demonstrated using either free oligonucleotides or synthetic DNA substrates bearing the IN recognition sequence motif (see, e.g., Katz et al., supra; and Bushman and Craigie, J. Virol. 64: 5648, 1990). Synthetic DNA substrates can be readily engineered by inserting a unique restriction enzyme site (typically *NdeI*), flanked by the appropriate IN recognition sequences, into a plasmid vector. Digestion of the vector with *NdeI* yields a DNA substrate with 3' recessed ends preceded by the highly conserved 5'-CA-OH dinucleotide and the remainder of the appropriate IN recognition motif, which resembles the processed ends of the pre-integrative viral DNA with which IN interacts to mediate integration. This approach has been successfully used to demonstrate in vitro integration of such linearized DNA substrates into double-stranded DNA in vitro by purified recombinant avian retrovirus IN (Katz et al., supra),

-34-

MoMLV IN (Bushman and Craigie, J. Virol., supra) and HIV IN (Bushman et al., Science, supra). The IN recognition sequences used on the termini of the substrate DNA in these experiments were short (10-30 base pairs), demonstrating that heterologous DNA substrates with short terminal IN sequence recognition motifs can be integrated into double-stranded DNA by IN. Moreover, these experiments document successful integration of both ends of the DNA substrate into the target DNA, as opposed to the oligonucleotide integration experiments which assay only for integration of a single end of the substrate DNA. These experiments document that linear DNA substrates bearing short terminal IN recognition motifs can be integrated into double-stranded DNA in vitro by purified IN protein. Thus, the foregoing experiments provide further evidence of the utility of the present invention, in that IN components can be included in the GDFP and can act in concert with terminal IN recognition sequence motifs present on the (substrate) tNA to mediate efficient integration. Recombinant fusions between integrase and heterologous proteins have previously been constructed, expressed and shown to retain integrase enzymatic activity (see, e.g., Vink et al., J. Virol., 68:1468-1474, 1994). Moreover, Bushman (PNAS 91:9233, 1994) has shown that recombinant fusions can be made between integrase and a sequence-specific DNA binding protein, and that such fusions retain integrase activity and sequence-specific DNA binding.

Thus, for example, by including an RI component derived from an integrase protein in the GDD, and using a tNA bearing the IN recognition sites, the GDFP can be co-introduced with the targeted nucleic acid (tNA) bearing the integration recognition motif and thus achieve integration of the tNA into a replicon of the target cell. This system would also allow, in conjunction with an appropriate binding domain in the NBD, for the association of the RI component with the free ends of the tNA. This would be advantageous since the IN proteins of retroviruses function to mediate integration at the free ends of pre-integrative viral DNA. In the present invention, this can be achieved by utilizing a Type-I GDFP in conjunction with a linearized tNA containing the cognate recognition sequence ("CRS") for the NBD at (or in close proximity to) the ends of the tNA bearing the terminal IN sequence recognition motif (preferably less than 500 nucleotides from the IN sequence, more preferably less than 200 nucleotides, most preferably less than 50 nucleotides). To generate the tNA for

-35-

gene delivery, the tNA can be constructed, for example, with a unique *NdeI* site between the two IN recognition motifs, flanked by the cognate recognition sequence of the NBD. Digestion with *NdeI* would then generate a linear DNA molecule with 3' recessed ends preceded by the 5'CA-OH dinucleotide, the remainder of the IN sequence recognition motif, and the CRS for the NBD, in that order. In this way, the terminal IN recognition sequences would be closely linked to the cognate recognition sequence for the NBD. Typically, the *NdeI* site/IN recognition sequence/CRS cassette would be inserted into the plasmid backbone of the vector. However, it is possible to construct a tNA devoid of any extraneous or undesirable sequences; for example, a tNA devoid of any bacterial plasmid sequence can be generated by flanking each end of the mammalian expression cassette in the tNA with a CRS, followed by one half of the IN recognition sequence, followed by an *NdeI* site. Digestion by *NdeI* would then generate a linear tNA DNA fragment, which could be readily purified from the plasmid backbone fragment, having on each end the IN recognition sequence and the CRS.

Removal of plasmid backbone sequences may be desirable to achieve optimal gene regulation in the transduced cells. Binding of the GDFP would then locate the RI component containing the IN region in close proximity to the sites on the tNA with which it can mediate efficient integration. An analogous strategy can be used with the AAV Rep protein and viral ITRs (see, e.g., Owens et al., J. Virol. 67:997-1005 (1993), and the review by Carter, B.J., Current Opinion Biotech. 3:533-539 (1992) and publications reviewed therein). Additionally, other recombinase systems such as the bacteriophage P1 cre recombinase, the yeast FLP recombinase, the yeast SR1-derived R recombinase and the Ty1 integrase (see, e.g., Kilby et al., Trend Genet., 9:413-421, 1993; Moore and Garfinkel, PNAS 91:1843-1847, 1994) can be used in the context of the present invention using fusions with appropriate NBDs, cis-acting recombinase recognition sequences and CRS elements, in an analogous fashion to that described above for the integrase fusions.

Multimerizing the RI domain may be required for optimizing integration activity, especially in situations in which the protein from which the RI domain is derived functions in multimeric form. Thus, for example, many native retroviral IN proteins are dimeric or multimeric in form (see, e.g., Jones et al., J. Biol Chem 23: 16037, 1992; reviewed in Skalka, Gene 135:175, 1993). Multimerization of the IN

-36-

domain can be conveniently achieved by, for example: (a) constructing tandem repeats of the IN component in the GDFP, preferably separated by a flexon; (b) dimerizing the GDFP by insertion of a protein dimerization motif, e.g., a leucine zipper motif (see, for example, Hu et al., 1990, Science 250: 1400); (c) adding free IN protein to an IN-
5 containing GDFP (since IN proteins have a natural tendency to multimerize); or (d) multimerizing the CRS in the tNA such that multiple GDFP molecules bind to each end of the tNA. Combinations of the above strategies can also be used. This would result in further multimerization of the IN component and thus a more active integration complex.

10 Another strategy to achieve multimerization of the RI domains, and also to achieve more efficient concerted integration of the tNA (i.e. having both ends of the tNA integrate into the replicon), would be to engineer the system to bring the free ends of the linear tNA together. This can be achieved in a number of ways in the context of the present invention. First, the GDFP monomers can be designed in such a way that
15 they self-dimerize, using a leucine zipper or other motif as described above. Thus, dimerization of the GDFP bound to the tNA would result in close apposition of the free ends of the tNA, since the CRS is located near these termini. A second approach would be to use a second, separate DNA binding protein with additional cognate recognition sequences present near the termini of the tNA to bring the ends of the tNA
20 together (for this purpose any of the DNA binding domains alluded to above could be used in dimeric form together with a tNA having the appropriate cognate recognition sequence to associate the free DNA ends). Such strategies would bring the free ends of the tNA in close apposition to one another and thus may further enhance the frequency of concerted integration. Several approaches can be used to avoid any potential
25 problem that may arise from GDFP/tNA complex auto-integration (i.e. integration of the ends of the tNA molecule into itself), or cross-integration of one tNA molecule into another. Although complexing of the GDFP with the tNA could be done at 4°C, thus reducing IN enzymatic activity, placing the complex onto cells at higher temperatures could lead to such unwanted integration events. A preferred approach is to use a
30 conditional RI moiety. Such conditional RI moieties can be dependent on chemical or protein co-factors, or can be mutants that are conditional for full activity dependent on temperature or other variables, such as the presence or absence of inhibitors or co-

-37-

factors. For example, in the case of IN, temperature-sensitive (*ts*) IN mutants have been made that are active only at certain temperatures (see, for example, Schwartzberg et al., Virology 192:673, 1993). The use of a *ts* RI component would allow exposure to and uptake of the complex by cells to be done at the non-permissive temperature (such that the RI component would not be active), followed by switching to the permissive temperature once the complex was taken up into the nucleus, allowing the RI component to be active in the context of the host cell replicon and thus accomplish the desired integration.

Thus, inclusion of an RI component in the GDFP can be used to enhance frequencies of integration. The GDD can consist of an RI component alone, or it can in addition comprise one or more of the other components discussed above. Where the RI component is the sole component in the GDD, the NBD would function to associate the RI component more stably and/or more specifically with the free ends of the tNA than is possible through, for example, use of the recombinant native IN protein alone. By virtue of the NBD binding, the RI moiety is more tightly associated with the tNA termini during the transfection process and can mediate integration into the host cell replicon. Where the RI component is the sole component in the GDD, the tNA/GDFP complex can be delivered by any of the standard means of transfection, such as lipofection, electroporation, etc., and the resulting cells would have an enhanced frequency of stable gene delivery as a consequence of enhanced integration of the tNA. Alternatively, the complex can be delivered by other non-viral means, including for example the use of self-assembling systems such as viral capsid proteins. Experimental evidence has confirmed that viral capsid proteins can be used to introduce DNA into mammalian cells (see, e.g., Forstova et al., Hum. Gene Ther., 6:297-306, 1995).

In certain cases, such as the retroviral IN or AAV Rep proteins, components of the GDD can also function as effective NBD components and thus fulfill a dual function in the GDFP by virtue of their ability to bind nucleic acid (see, e.g., Krongstad & Champoux, J. Virol. 64:2796-2801, 1990; Owens et al., J. Virol. 67:997-1005 (1993); and the review by Carter, B.J., Current Opinion Biotech. 3:533-539 (1992) and publications reviewed therein).

2. The Targeted nucleic acid (tNA)

The targeted nucleic acid (tNA) is a polynucleotide, or analog thereof, to be delivered to a target cell. Thus, targeted nucleic acids include, for example, oligonucleotides and longer polymers of DNA, RNA or analogs thereof, in double-stranded or single-stranded form. The tNA may be either circular, supercoiled or linear. A preferred example of a targeted nucleic acid is a DNA expression vector comprising a gene (or genes) of interest operably linked to a transcriptional control region (or regions) and a cognate recognition sequence capable of being bound by the NBD domain of the GDFP. The transcriptional control region may be selected so as to be specifically activated in the desired target cells, or to be responsive to specific cellular or other stimuli.

Targeted nucleic acids may also include, for example, positive and/or negative selectable markers; thereby allowing the selection for and/or against cells stably expressing the selectable marker, either in vitro or in vivo.

Use of the present invention to deliver RNA would enable the introduction of RNA decoys (Sullenger et al., Cell 63:601-608, 1990); ribozymes (Young et al., Cell 67:1007-1019, 1991); and antisense nucleic acids (Vickers et al., Nucleic Acid Res., 19:3359-3368), for example.

In Type-I GDFPs, the targeted nucleic acids are recognized and bound by the GDFP by virtue of specific cognate recognition sequences to which the nucleic acid binding domain (NBD) of the Type-I GDFP binds. Both DNA and RNA binding domains have been isolated from proteins that bind to particular nucleic acids in a sequence-specific fashion. Inclusion of such a cognate recognition sequence in the targeted nucleic acid allows for specific binding of the GDFP to the tNA. Recognition sites for many nucleic acid binding proteins have been identified (see, for example, Mitchell & Tjian, 1989; and other references herein).

Binding of sequence-specific binding proteins to DNA tends to be more avid when the recognition sequence motif is multimerized (see, e.g., Hochschild and Ptashne, Cell 44:681-687, 1986). Accordingly, the cognate recognition sequences may be multimerized in the targeted nucleic acids so as to enhance the binding affinity or selectivity of a GDFP for its cognate tNA. This could also have other advantages, such as increasing the effective amount of the GDFP bound to the tNA, or promoting

-39-

compaction/condensation of the tNA by sequence-specific or sequence-non-specific NBD components.

Typically, but not necessarily, the cognate recognition sequences in expression vectors will be placed in the plasmid backbone of the vector. This also applies to other
5 cis-acting sequences that are needed in the tNA to facilitate gene delivery. However, it may be desirable to remove plasmid backbone sequences from the DNA to be transferred. In this case, the expression cassette can be conveniently flanked by restriction enzyme sites, such that restriction enzyme digestion separates the backbone from the mammalian expression cassette. The expression cassette can then be purified
10 away from the plasmid backbone for use in transduction experiments. Clearly in this case the CRS would need to be located on the fragment bearing the expression cassette. It is also possible, of course, to construct the GDFP so as to bind to more than one tNA.

As discussed above, the tNA can also be bound to the GDFP via sequence-non-specific interactions in addition to sequence-specific interactions. In a Type-I GDFP,
15 such sequence-non-specific interactions can be mediated by auxiliary components derived from sequence-non-specific binding proteins, as discussed above. Such auxiliary non-specific binding components can also serve to compact or otherwise reconfigure the targeted nucleic acid; see, supra.

20 The targeted nucleic acids can also include, for example, non-expressed DNA, such as sequences homologous to sequences present in a target cell replicon, that can thereby mediate homologous recombination. This can be used to facilitate the stable integration of the targeted nucleic acid, or a desired portion thereof, into a specific site in a replicon present in the target cell, such as a specific site in a cellular chromosome.
25 This may be useful, for example, to achieve a desired level of expression of the tNA by integration at a desired chromosomal site. Homologous recombination can also be used to alter a specific DNA sequence in a target cell replicon (see, e.g., Thomas & Capecchi, Cell 51:503-512, 1987).

30 For longer tNA sequences, or where the tNA uptake mechanism (whether part of the GDFP or not) is known or suspected to be sensitive to the size, form or charge of nucleic acids and/or complexes to be delivered, such as mechanisms involving endocytosis, it may be desirable to condense and/or charge neutralize the tNA. This

-40-

can be achieved by mixing the tNA with any of a number of proteins or other agents (collectively referred to as "compacting agents") that can condense and/or charge neutralize nucleic acids. Compacting agents include, for example, histones (see, e.g., von Holt, *Bioassays* 3:120-124, 1986; and Rhodes, *Nucleic Acid Res.*, 6:1805-1816, 1979); or polypeptides derived therefrom (Rodriguez et al., *Biophys. Chem.*, 39:145-152, 1991); as well as the non-histone high mobility group proteins. Poly-L-lysine or other polybasic amino-acids can also be used as compacting agents (see, e.g., Li et al., *Biochemistry*, 12:1763-1772, 1973; and Weiskopf and Li, *Biopolymers* 16:669-684, 1977). Similarly, other polycationic polymers such as polyamines, for example spermine and spermidine, and cationic lipid-containing polymers can also be used to condense and/or charge neutralize nucleic acid (see, e.g., Feuerstein et al., *J. Cell. Biochem.*, 46:37-47, 1991; and Behr, *Bioconj. Chem.*, 5:382, 1994). Retroviral nucleocapsid proteins can fulfill a similar role (see, e.g., Gelfand et al. *J. Biol. Chem.*, 268:18450-18456, 1993).

Alternatively, compacting agents can be incorporated as an additional component of the GDFP. Also, some sequence specific binding proteins, such as GAL4, which exhibit a range of binding affinities to different cognate nucleic acid sequences may also be used in this capacity, and in this regard would function as an NBD with both nucleic acid binding and compaction properties.

Compacting agents might also be incorporated as mediators of indirect binding between the tNA and the NBD domain of the GDFP (for example, the NBD domain can be bound to the compacting agent and the compacting agent bound to the tNA).

Assembly of GDFPs

Preferably, the GDFP is prepared as a single polypeptide fusion protein generated by recombinant DNA methodology. To generate such a GDFP, sequences encoding the desired components of the GDFP are assembled and fragments ligated into an expression vector. Sequences encoding the various components may be assembled from other vectors encoding the desired protein sequence, from PCR-generated fragments using cellular or viral nucleic acid as template nucleic acid, or by assembly of synthetic oligonucleotides encoding the desired sequence. However, all nucleic acid sequences encoding such a preferred GDFP should preferably be assembled by in-frame

-41-

fusions of coding sequences. Flexons, described above, can be included between various components and domains in order to enhance the ability of the individual components to adopt configurations relatively independently of each other.

Although a Type-I GDFP is preferably assembled and expressed as a single polypeptide chain, one or more of its domains or components may be produced as a separate chain that is subsequently linked to the GDFP by, e.g., disulfide bonds, or chemical conjugation. It is also feasible to prepare complexes in which domains such as the NBD and the GDD or their components are physically associated by other than recombinant means, either directly or indirectly, for example, by virtue of non-covalent interactions, or via co-localization on a proteinaceous or lipid surface.

The GDFP may be expressed either in vitro, or in a prokaryotic or eukaryotic host cell, and can be purified to the extent necessary. An alternative to the expression of GDFPs in a host cell is synthesis in vitro. This may be advantageous in circumstances in which high levels of expression of a GDFP might interfere with the host cell's metabolism; and can be accomplished using any of a variety of cell-free transcription/translation systems that are known in the art. GDFPs can also be prepared synthetically. It will likely be desirable for the GDFP to possess a component or sequence that can facilitate the detection and/or purification of the GDFP. Such a component may be the same as or different from one of the various components described above.

Many approaches of expressing and purifying recombinant proteins are known to those skilled in the art, and kits for recombinant protein expression and purification are available from several commercial manufacturers of molecular biology products. Typically, an increased level of purity of the GDFP will be desirable. However, because of the specificity of the GDFP for nucleic acid binding, the degree of purification need not necessarily be extensive. The GDFPs of the present invention may be sterilized by simple filtration through a 0.22 or 0.45u filter so as to avoid microbial contamination of the target cells.

Since the domains of the GDFP can be assembled in modular fashion in an expression vector, its construction by recombinant DNA methodology allows the GDD to consist of one or many components. Such components may have complementing activity in mediating or enhancing gene delivery, or they may have closely related

-42-

functions. In essence, the gene delivery domain can be viewed as possessing any function that mediates or enhances the efficiency of delivery of the tNA bound to the GDFP.

5 Other Variations of GDFPs

Other variations will be apparent to those of skill in the art. For example, the GDFP may itself be multimerized. Multimerization may be advantageous to increase avidity of binding of either the NBD or the GDD. A given tNA molecule may also contain multiple distinct cognate recognition sequences, binding different Type I
10 GDFPs with distinct functions, or the tNA may be bound with a mixture of Type I and Type II GDFPs. Additionally, certain components of the GDD, such as IN proteins, may require dimerization for optimal activity. Dimerization of the GDFP may be obtained by including, for example, a leucine zipper motif in the GDFP. Such motifs are common in DNA binding proteins and are responsible for their dimerization
15 (Kouzarides & Ziff, 1989). Leucine zippers can be inserted into DNA binding proteins and cause them to dimerize (Sellers and Struhl, Nature 341:74-76, 1989). Multimerization of GDFPs can also be achieved, for example, by creating a recombinant fusion protein that contains two or more GDFPs. Preferably such multimerized GDFPs are separated by flexons, as described herein. Other
20 oligomerization motifs from dimeric or multimeric proteins can similarly be employed.

Illustrations of Type-II Gene Delivery Fusion Proteins

Type-II GDFPs do not bind targeted nucleic acids in a sequence-specific manner
25 because the nucleic acid binding components of Type-II GDFPs are all derived from nucleic acid binding proteins that are non-sequence-specific in their binding to nucleic acid.

Nucleic Acid Binding Domains of Type-II GDFPs

30 The nucleic acid binding domains (NBDs) of Type-II GDFPs comprise binding components that are derived from non-sequence-specific nucleic acid binding proteins, recombinantly fused to a gene delivery domain (GDD) as described above.

-43-

A number of non-sequence-specific nucleic acid binding proteins have been identified and characterized, including, for example, histones or polypeptides derived therefrom (see, e.g., von Holt, Bioassays 3:120-124, 1986; Rhodes, Nucleic Acid Res., 6:1805-1816, 1979; and Rodriguez et al., Biophys. Chem., 39:145-152, 1991);
5 retroviral nucleocapsid proteins (see, e.g., Gelfand et al. J. Biol. Chem., 268:18450-18456, 1993); proteins such as nucleolin (Erard et al., Eur. J. Biochem. 191:19-26, 1990); avidin (Pardridge & Boado, F.E.B.S. Lett. 288:30-32, 1991); and polybasic polypeptide sequences such as poly-L-lysine (Li et al., Biochemistry, 12:1763-1772 1973; Weiskopf and Li, Biopolymers 16:669-684, 1977).

10 For the reasons discussed herein, all of the GDFPs of the present invention are preferably produced as recombinant fusion proteins. However, the recombinant expression, in a host cell, of non-sequence-specific nucleic acid binding components in Type-II GDFPs (as well as in Type-I GDFPs that incorporate sequence-non-specific nucleic acid binding components) may be hindered by interference of the expressed
15 proteins with host cell nucleic acids. In such situations, the GDFPs can be readily synthesized in vitro using any of a variety of cell-free transcription/translation systems that are known in the art.

Gene Delivery Domains of Type-II GDFPs

20 The various possible sources of components making up the gene delivery domains of Type-II GDFPs are essentially the same as described above for Type-I GDFPs (although, by definition, Type-II GDFPs would not include sequence-specific binding components such as the sequence-specific integrase components described above for Type-I GDFPs).

25

Targeted Nucleic Acids for Use with Type-II GDFPs

The targeted nucleic acids to be combined with Type-II GDFPs are as described above except that they need not contain specific recognition sequences since the Type-II GDFPs bind nucleic acids via non-specific interactions.

30

-44-

Assembly of Type-II GDFPs

The assembly of Type-II GDFPs is preferably via the synthesis of recombinant fusion proteins (see the description above regarding assembly of Type-I GDFPs).

5 Using GDFPs of the Present Invention

Thus, the GDFPs of the present invention can be used for in vitro or in vivo gene delivery. For therapeutic applications, target cells can be transduced ex vivo and returned to a patient, or, given the biochemical nature of the tNA/GDFP complex, cells can be treated directly in vivo. For such in vivo therapy, the complexes can be
10 formulated for a variety of modes of administration, including systemic and topical or localized administration. Techniques and formulations may be found, for example, in Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, PA. (latest edition). The tNA/GDFP complex may be combined with a carrier such as a diluent or excipient which may include, for example, fillers, extenders, wetting agents,
15 disintegrants, surface-active agents, or lubricants, depending on the nature of the mode of administration and the dosage forms. The nature of the mode of administration will depend, for example, on the location of the desired target cells. For in vivo administration, injection is preferred, including intramuscular, intratumoral, intravenous, intra-arterial (including delivery by use of double balloon catheters),
20 intraperitoneal, and subcutaneous. Delivery to lung tissue can be accomplished by, e.g., aerosolization. For injection, the complexes of the invention are formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the complexes may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are
25 also included. Systemic administration can also be by transmucosal or transdermal means, or the compounds can be administered orally. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. For topical administration, the complexes of the invention may be formulated into ointments, salves, gels or creams, as is generally known in the art.

30 The GDFP approach can thus be used to target any cell, in vitro, ex vivo or in vivo, the only requirement being that the target cells have binding sites for the GDFP on their surface. The present invention will thus be useful for many gene therapy

-45-

applications. As an illustrative example, the target cells that could be used in the context of the present invention include lymphohematopoietic cells. These include: (i) stem cells, which have many potential applications in gene therapy, including correction of hereditary disorders such as Gaucher disease and hemoglobinopathies, as well as genetic modification with intracellular vaccines against HIV such as decoys or dominant-negative proteins; and (ii) lymphocytes, which would allow genetic modification of effector T cells such as CTLs for use in human therapy with genes of interest such as suicide genes and regulated promoter cassettes. Also included for use in the context of the present invention are cells of the cardiovascular system which line blood vessels including endothelial cells and vascular smooth muscle cells, which could be genetically modified to inhibit atherosclerosis or restenosis following angioplasty. Similarly, the present invention could be used to introduce genes into airway epithelial cells, such as the CFTR gene to correct cystic fibrosis. The present invention could also be used to transduce tumor cells and thereby genetically modify them to express suicide genes for tumor elimination or produce cytokines or express immunostimulatory molecules for use as a tumor vaccine in cancer patients. Another illustrative application of the present invention is delivery of DNA or RNA to antigen presenting cells (APCs). This could be useful, for example, to allow expression of specific (tNA-encoded) antigens by an APC, thereby allowing the APC to stimulate an antigen-specific immune response, such as a CTL response. Such an approach can be used in vitro, by transduction of APCs with a GDFP/tNA complex thereby allowing antigen presentation for the stimulation and generation of CTLs in vitro, or in vivo delivery can be used, to allow such antigen presentation in vivo. Direct delivery of RNA to APCs using the present invention may be especially desirable for situations in which antigens are encoded by transcripts that require special conditions for intracellular transport or processing that may not happen efficiently in the APC. An illustrative example would be rev-dependent RNAs of HIV (such as HIV gag). Transduction of APCs with RNA in the context of the present invention can thus be used, for example, to circumvent the need for nuclear export of rev-dependent RNAs. Additionally, the present invention could be used to introduce genes into hepatocytes of the liver to correct genetic defects such as familial hypercholesterolemia, hemophilia and other metabolic disorders, or to produce recombinant products for systemic delivery.

-46-

Similarly, fibroblasts or connective tissue cells could be modified to secrete cytokines or soluble enzymes for immunomodulatory purposes or to correct a metabolic deficiency. These tissue targets and diseases, together with others are more fully described in Scriver et al., Eds., 'The Metabolic Basis of Inherited Disease', 6th Ed., McGraw-Hill, 1989, and in Miller, A.D., Blood 76:271-278, 1990. The present invention is particularly useful in cases in which genes of interest cannot be transferred by commonly used viral vectors, or in which the target cells are not infectable by viral approaches (see, e.g., Israel & Kaufman, Blood, 75:1074-1080, 1990; Shimotohno & Temin, Nature 299:265-268, 1982; Stead et al., Blood, 71:742-747, 1988; and Bodine et al., Blood, 82:1975-1980, 1993).

The GDFP approach of the present invention can be used as a generically useful method for gene transduction of cells, and could be provided as a laboratory kit for gene transduction for use with, e.g., insect, avian, mammalian, or other higher eukaryotic cells.

The transfer of genes in the present invention can also be facilitated by other biochemicals known to enhance the uptake of nucleic acid by cells (see, e.g., Kawai & Nishizawa, Mol. Cell. Biol. 4:1172-1174, 1984; Behr et al., P.N.A.S. 86:6982-6986, 1989; Rose et al., P.N.A.S. Biotechniques 10:520-525, 1991; Pardridge & Boado, F.E.B.S. Lett. 288:30-32, 1991; Legendre & Szoka, P.N.A.S. 90:893-897, 1993; Haensler & Szoka, Bioconj. Chem. 4:372-379, 1993). These and other techniques for use in the context of the present invention can be used under conditions (for incubation etc.) as described in the art (see, e.g., Kriegler, M. 1990 (ed.), "Gene Transfer and Expression, a Laboratory Manual," (1990)). In the case of GDFPs comprising pH-dependent M-D components, such as the TM protein of diphtheria toxin (see, e.g., Choe et al. (1992) Nature 357:216-222), entry of the GDFP/tNA complex into the cell can be conveniently achieved by simply reducing the pH of the incubation medium during transduction.

The examples presented below are provided as a further guide to the practitioner of ordinary skill in the art, and are not to be construed as limiting the invention in any way.

-47-

Example iPreparation of a Nucleic Acid Binding Domain (NBD)From the Yeast GAL4 Protein

5 The DNA binding domain of GAL4, amino acids 1-147 (Laughon and Gesteland, Molecular and Cellular Biology 4:260-267, 1984; Ma and Ptashne, Cell 48:847-853, 1987; and Carey et al., J. Mol. Biol. 209:423-432, 1989), was amplified by PCR from S. cerevisiae (ATCC 60248) using the following amplimers.

The amplimer for the 5' end of GAL4 was as follows:

5' GCGC ACTAGT GCCACC ATG AAG CTA CTG TCT TCT ATC G 3'.

10 The GAL4 coding region is underlined. This amplimer created a SpeI site (ACTAGT) for cloning into pBluescript (Stratagene) which allowed for subsequent transcription by T3 RNA polymerase. The amplimer also included a consensus sequence (GCCACC) for efficient protein translation located upstream of the initiator methionine (Kozak et al., Nucl. Acids Res. 15:3374, 1987).

15 The amplimer for the 3' end of the GAL4 NH2-terminus (up to amino acid 147) was as follows:

5' GCGC GGTACC TCCGGA TAC AGT CAA CTG TCT TTG ACC 3'.

20 The GAL4 coding region is underlined. This amplimer created a 3' Asp718 site (GGTACC) for cloning into pBluescript as noted above. The amplimer also included a BspE1 site (TCCGGA) to allow for an in-frame fusion with an oligomer encoding a flexible peptide sequence (see below).

25 The GAL4 fragment was amplified by 30 cycles of PCR directly from a colony of S. cerevisiae. The product was digested with SpeI and Asp718 and ligated between the SpeI and Asp718 sites located in the polylinker region of pBluescript. The construct was transformed into the DH10B strain of E. coli by electroporation, and a colony containing the GAL4 fragment was identified by restriction enzyme analysis.

The resulting plasmid, designated pT3gGAL4, is shown in Fig. 2A.

-48-

Example 2Preparation of a Gene Delivery DomainFrom the Human IL-2 Protein

5 A DNA fragment encoding mature soluble human IL-2 (amino acids 21-133) was amplified by PCR from a full-length human IL-2 cDNA (Taniguchi et al., Nature 302:305-310, 1983), using the following amplimers.

The amplimer for the 5' end of mature human IL-2 was as follows:

5' GCGC ACTAGT GCCACC ATG GCG CCT ACT TCA AGT TCT ACA AAG
AAA AC 3'.

10 The IL-2 coding region is underlined. This amplimer created a SpeI site (ACTAGT) for cloning into pBluescript, and inserted an initiator methionine immediately upstream of amino acid 21 of IL-2. The amplimer also contained a consensus sequence (GCCACC) for efficient translation upstream of the inserted methionine, as noted above. A NarI site (GGCGCC) was also included which allowed
15 for a subsequent in-frame fusion with a linker sequence which separated the GAL4 and IL-2 domains in the GAL4/IL-2 construct (see below).

The amplimer for the 3' end of mature human IL-2 was as follows:

5' GCGC GGTACC TCA AGT CAG AGT ACT GAT GAT GCT TTG ACA AAA
GGT AAT C 3'.

20 This amplimer created an Asp718 site (GGTACC) for cloning into pBluescript, and also retained the wild-type termination codon for human IL-2. A ScaI site (AGTACT) at the 3' end of the IL-2 coding region was also created by this amplimer without introducing amino acid changes. The DNA fragment encoding the mature human IL-2 protein was amplified by 30 cycles of PCR from the full-length human
25 IL-2 cDNA referred to above. The product was digested with SpeI and Asp718, ligated into pBluescript, and transformed into DH10B cells as described above. A colony harboring an appropriate construct was identified by restriction enzyme analysis.

Sequencing of a plasmid derived from one colony revealed that an alteration (loss of a single base resulting in a frame-shift near the terminus of IL-2) had occurred
30 within the 3' amplimer during PCR cloning - thereby generating an IL-2 mutein. Specifically, the first T after the Sca I site (in the first GAT triplet) was removed, causing a frame-shift that also generated a premature termination codon. As a result,

-49-

the 5 amino acids normally present at the terminus were replaced by 3 different amino acids. This plasmid, referred to as "pT3matIL-2m" (shown generically in Figure 2A as pT3matIL-2), was used to create a gene delivery fusion protein as described in Example 3. Despite the variation in the IL-2 domain, a GDFP based on this IL-2

5 mucin exhibited IL-2 bioactivity, as described below.

A second colony contained a plasmid designated "pT3matIL-2" (as shown generically in Figure 2A) that contained the expected wild-type IL-2 sequence. Plasmid pT3matIL-2 was used to create two GDFPs as described in Examples 3 and 4.

10

Example 3

Construction of Plasmids Encoding a Gene Delivery

Fusion Protein (GDFP) Having a GDD and an NBD

Separated by a Flexon

A DNA fragment encoding the nucleic acid binding domain (NBD) derived from

15 GAL4 was isolated from pT3gGAL4 (Example 1) by digesting with Spe1 and BspE1.

A DNA fragment encoding the gene delivery domain (GDD) derived from a human IL-2 mucin was isolated from pT3matIL-2m (Example 2) by digesting with Nar1 and Asp718. The following oligomer pair encoding the flexon sequence (GlyGlyGlyGlySer), was annealed creating a 5' BspE1 over-hang (CCGGA) and a 3'

20 Nar1 over-hang (CGCC):

5' CCGGA GGC GGT GGA TCC GGT GGT GGA GGC AGT GGA GGA GGT GGC
TCGG 3';

5' CGC CGA GCC ACC TCC TCC ACT GCC TCC ACC ACC GGA TCC ACC
GCC T 3'.

25

The NBD and GDD fragments and the annealed oligomer were ligated into pBluescript between the Spe1 and Asp718 sites, and transformed into DH10B cells as described above. A colony harboring a construct that contained all three fragments was identified by its ability to hybridize to both GAL4 and IL-2 [³²P]-labeled fragments, and by restriction enzyme analysis.

30

In the resulting plasmid, designated "pT3GAL4/IL-2m" (shown generically in Fig. 2A as pT3GAL4/IL-2), the sequence encoding the GDFP was inserted into pBluescript in an orientation which allowed for sense RNA transcripts to be synthesized

-50-

with T3 RNA polymerase. The resulting RNA, when translated, incorporated both the DNA binding domain of the yeast GAL4 protein and the mature form of the human IL-2 mutein, in that order, separated by a flexible amino acid linker.

5 A second plasmid, designated pT3GAL4/IL-2 (as shown in Fig.2A), was constructed exactly as described for pT3GAL4/IL-2m, except that the DNA fragment encoding the gene delivery domain (GDD) derived from human IL-2 was isolated from pT3matIL-2 (Example 2).

Example 4

10 Construction of a Third Plasmid Encoding a Gene Delivery
Fusion Protein (GDFP) Having a GDD and an NBD
Separated by a Flexon

Another expression vector encoding a GDFP derived from IL-2 and GAL4 was constructed as follows.

15 The DNA binding domain of GAL4, amino acids 1-147 (Carey, et al., supra), was amplified by 30 cycles of PCR from pT3gGAL4 using the following amplimers.

The amplimer for the 5' end of GAL4 was as follows:

5' GCGC GGATCC ATG AAG CTA CTG TCT TCT ATC G 3'.

20 This amplimer created a BamH1 site (GGATCC) immediately upstream of Met¹ to allow for an in-frame fusion with a flexible peptide sequence in front of GAL4 (see below).

The amplimer for the 3' end of the GAL4 NH₂-terminus (up to amino acid 147) was as follows:

5' GCGC GGTACC G CTA GCT TAC AGT CAA CTG TCT TTG ACC 3'.

25 This amplimer created an Asp718 site (GGTACC) for cloning into pBluescript and also included an engineered termination codon (CTA) at the C-terminus of the DNA binding domain of GAL4.

30 To construct pT3IL-2/GAL4, the GAL4 PCR product was digested with BamH1 and Asp718. A DNA fragment encoding human IL-2 was isolated from pT3matIL-2 (see Example 2) by digesting with Spe1 and Sca1. The following oligomer pair encoding the amino acid sequence (GlyGlyGlyGlySer)₃ was annealed, creating a 5' Sca1 over-hang (ACT) and a 3' BamH1 over-hang (GATCC):

-51-

5' ACT CTG ACT GGA GGT GGG GGC TCT GGT GGC GGA GGT AGT GGA
GGA GGT G 3';
5' GA TCC ACC TCC TCC ACT ACC TCC GCC ACC AGA GCC CCC ACC TCC
AGT CAG AGT 3'.

- 5 The IL-2 and GAL4 fragments and oligomers were ligated into pBluescript between the SpeI and Asp718 sites and the construct was transformed into the DH10B strain of *E. coli* by electroporation. A colony containing all three fragments was identified by its ability to hybridize to both GAL4 and IL-2 [³²P]-labeled fragments, and by restriction enzyme analysis.
- 10 In the pT3IL-2/GAL4 construct, shown in Figure 2B, the GDFFP was inserted into pBluescript in an orientation which allowed for a sense RNA to be synthesized with T3 RNA polymerase. The resulting RNA, when translated, incorporated both the mature form of human IL-2, and the DNA binding domain of the yeast GAL4 protein, in that order, separated by a flexible amino acid linker.

15

Example 5

Expression of Gene Delivery Fusion Proteins

- Sense mRNA encoding the GAL4/IL-2m GDFFP construct (described in Example 3) was transcribed in vitro with T3 RNA polymerase from the pT3GAL4/IL-2 vector.
- 20 Briefly, pT3GAL4/IL-2m plasmid was linearized with Asp718 and this template was combined with a ribonucleotide mixture (rNTPs), RNA cap structure analog (m7Gppp), and T3 RNA polymerase in a HEPES-based buffer (Promega "RiboMAX"). After incubation at 37 degrees C, the DNA template was digested with RNase-free DNase (Promega), and the synthesized mRNA was separated from unincorporated
- 25 rNTPs by chromatography through a G25 Sephadex spin column (Boehringer Mannheim), precipitated with EtOH, and quantitated by OD₂₆₀.

- The resultant mRNA was translated in a cell-free rabbit reticulocyte lysate system. mRNA was added to a translation mixture of reticulocyte lysate, RNasin, and complete amino acids (Promega). Translation was allowed to proceed for 1 to 2 hr. at
- 30 30 degrees C, after which lysates were stored at -70 degrees C. The integrity and molecular weight of the fusion protein was assessed by including [³⁵S]-labeled methionine (Amersham) in the translation mix, and visualizing the product by

-52-

polyacrylamide gel electrophoresis under denaturing conditions. Fig. 3, lane 2, shows the [³⁵S]-labeled GAL4/IL-2m translation product as resolved on a 14% acrylamide gel. The position of the GAL4/IL-2m GDFP translation product agreed with the predicted MW of 33kD. Molecular weight markers are shown in Fig. 3, lane 1, and a negative control is shown in lane 3.

Sense mRNAs encoding the GAL4/IL-2 GDFP construct (described in Example 3) and the IL-2/GAL4 GDFP construct (described in Example 4) were transcribed in vitro with T3 RNA polymerase from the pT3GAL4/IL-2 and pT3IL-2/GAL4 vectors, respectively, exactly as described above. Figure 6, lanes 3 and 4, show the [³⁵S]-labeled IL-2/GAL4 and GAL4/IL-2 translation products as resolved on a 4-20% gradient acrylamide gel. The positions of the IL2/GAL4 and GAL4/IL-2 GDFP translation products agreed with the predicted MWs of 33.3kD and 33.2kD, respectively. Molecular weight markers are shown in Fig. 6, lane 1, and a luciferase control is shown in lane 2.

Example 6

Sequence-Specific DNA Binding Activity of GDFPs

The ability of the GDFPs of Example 5 to engage in sequence-specific DNA binding was demonstrated by use of an electrophoretic mobility shift assay (EMSA) (Ausubel et al. (eds), "Current Protocols in Molecular Biology," (1987 and 1993)).

The target oligomer to which the GAL4 protein binds was:

5' TCGACGGAGTACTGTCCTCCGC 3'

3' GCCTCATGACAGGAGGCGAGCT 5'.

The following target oligomer is not bound by GAL4 and was used as a negative control:

5' TCGACTGAGTACTGTCCTCAGC 3'

3' GACTCATGACAGGAGTCGAGCT 5'.

The GAL4 target oligomer was end-labeled using [³²P]-dCTP (Amersham) and Klenow polymerase (New England BioLabs). The labeled oligomer was separated from unincorporated nucleotides by chromatography over a G25 spin column (Boehringer Mannheim), and quantified by scintillation counting. This oligomer was added to reactions containing a HEPES-based buffer, which included

-53-

poly(dI-dC)•poly(dI-dC) (Pharmacia), ZnCl₂, glycerol and BSA (Carey, et al., J. Mol. Biol. 209:423-432, 1989; Chasman, et al., Mol. and Cell. Biol. 9:4746-4749, 1989), and varying amounts of reticulocyte lysate containing either GAL4/IL-2m GDFP or IL-2. The reactions were electrophoresed on a 4.5% polyacrylamide/1% glycerol gel in 0.5x TBE. The gel was fixed in a methanol/acetic acid solution, dried, and analyzed on a Molecular Dynamics phosphorimager. The results shown in Figure 4 show decreasing amounts of input GAL4/IL-2m fusion protein (lanes 1-3) showing specific interaction of the GAL4/IL-2m GDFP with the labeled target oligomer. DNA size markers appear in lane 4. In lane 5, the GAL4/IL-2m GDFP was incubated with labeled target oligomer, as in lane 1, but excess unlabeled target oligomer was also included and competed with the labeled target oligomer for binding to the GAL4/IL-2m GDFP. In lane 6, the GAL4/IL-2m GDFP was incubated with labeled target oligomer, as in lane 1, but excess unlabeled non-binding oligomer was included, and showed lack of competition for binding of the GAL4/IL-2m GDFP to the labeled target oligomer. In lane 7, labeled target oligomer was incubated with lysate containing human IL-2, and showed no specific binding of the labeled target oligomer by either IL-2 or reticulocyte lysate components. The GAL4/IL-2m GDFP thus bound specifically to the cognate target sequence recognized by the GAL4 (NBD) domain of the GAL4/IL-2m GDFP.

The results in Figure 7 show sequence-specific binding of the GAL4 protein and the IL-2/GAL4 and GAL4/IL-2 GDFPs. Target oligomers, binding conditions, electrophoresis and gel treatment were exactly as described above, except that analysis was by autoradiography. The first four lanes contained decreasing amounts of input GAL4 protein, as indicated, showing specific interaction of GAL4 with the labeled target oligomer. The following lanes contained decreasing amounts of either input GAL4/IL-2 GDFP or input IL-2/GAL4 GDFP as indicated in Figure 7. The designation "+c" indicates that the GDFP was incubated with labeled target oligomer, as in previous lanes, but excess unlabeled target oligomer was also included in the reaction. The designation "+m" indicates that the GDFP was incubated with labeled target oligomer, as in previous lanes, but excess unlabeled non-binding oligomer was included in the reaction. The unlabeled target oligomer competed with the labeled target oligomer for binding to the GDFP while the non-binding oligomer showed lack

-54-

of competition, demonstrating specific binding of the GDFP to the GAL4 recognition sequence. In the lane designated "IL2," labeled target oligomer was incubated with lysate containing human IL-2, and showed no specific binding of the labeled target oligomer by either IL-2 or reticulocyte lysate components. The IL-2/GAL4 and
5 GAL4/IL-2 GDFPs thus bound specifically to the cognate target sequence recognized by the GAL4 (NBD) domain of the GDFPs.

Example 7

Cytokine Bioactivity of GDFPs

10 GAL4/IL-2m, GAL4/IL-2, and IL-2/GAL4 fusion proteins from in vitro translations (as in Example 5) were assayed for their IL-2 activities using the well-known CTLL bioassay. Cells were incubated with the GDFP, then pulsed with ³H-thymidine and incorporation of radioactivity into DNA was used as a measure of cellular proliferation, as described by Gillis et al., J. Immunol. 120:2027, 1978.

15 The results from the GAL4/IL-2m GDFP are shown in Figure 5. The IL-2 standard represents 1 ng/ml recombinant human IL-2 which was serially diluted 1:3 in the bioassay. The GAL4/IL-2m GDFP curve was generated using in vitro translated material starting with a 1:10 dilution of lysate. The bioassay shows retention of IL-2 biological activity by the GAL4/IL-2m GDFP.

20 The results from the GAL4/IL-2 and IL-2/GAL4 GDFPs are shown in Figure 8. The IL-2 standard was as described above. The GDFP curves were generated using in vitro translated material starting with a 1:50 dilution of lysate. The bioassay shows retention of IL-2 biological activity by the GAL4/IL-2 and IL-2/GAL4 GDFPs.

25

Example 8

Construction and Characterization of GDFPs Containing the Diphtheria Toxin Transmembrane Region.

The transmembrane (TM) domain of the Diphtheria toxin protein from C. diphtheriae, amino acids 205-378 (Choe, et al., Nature 357:216-222, 1992), is the
30 region responsible for endosomal release of the catalytic domain of the toxin into the cytoplasm of infected cells (Papini, et al., JBC 268:1567-1574, 1993, Madshus, JBC 269:17723-17729, 1994). This region has also been shown to be capable of cellular

-55-

membrane insertion in response to a mildly acidic environment (Moskaug, et al. JBC 263:2518-2525, 1988, McGill, et al., EMBO 8:2843-2848, 1989). To incorporate this domain into the pre-existing GDFPs, two DNA fragments encoding the transmembrane domain of the Diphtheria toxin protein were amplified by PCR from *C. diphtheriae* genomic DNA. The first fragment, termed "DT," encoded amino acids 205-378 and was amplified using the following amplimers.

The amplimer for the 5' end was as follows:

5' GTAGATCTGGTGGAGGTGGCTCCGGAGGAGGT GGATCC GAT TGG GAT
GTC ATA AGG GAT AA 3'

The amplimer for the 3' end was as follows:

5' CTTC AGATCT GGATCC T CCA CCG CCA CTA CCT CCA CCC CCG GGA
CGA TTA TAC GAA TTA TGA AC 3'

The toxin TM sequences are underlined. Both amplimers provided BamH1 sites near the termini for subsequent cloning of the PCR fragment into BamH1-digested pT3IL-2/GAL4 (Example 4).

The second fragment, termed "DAB," encoded amino acids 176-378, and provided additional residues at the amino terminus of the transmembrane region. Within the context of the intact toxin protein these additional sequences are involved in an enzymatic cleavage step which may be necessary for membrane fusogenic activity (Williams, et al., JBC 265:20673-20677, 1990, Ariansen, et al., Biochemistry 32:83-90, 1993). The "DAB" fragment was amplified using the following amplimers:

The amplimer for the 5' end was as follows:

5' GCG GGATCC GGT GGC GGA GGA AGT GAT GCG ATG TAT GAG TAT
ATG GCT C 3'

The amplimer for the 3' end was the same 3' amplimer used to PCR the above described "DT" fragment.

The toxin TM sequences are underlined. The resulting "DAB" PCR fragment was digested with BamH1, and cloned into the BamH1 sites of both pT3GAL4/IL-2 and pT3IL-2/GAL4 (Examples 3 and 4).

Thus, three triple-domain fusion plasmids were generated (pT3IL-2DTGAL4, pT3IL-2DABGAL4, and pT3GAL4DABIL-2), each containing a version of the diphtheria toxin transmembrane region as the middle domain. Each of the three GDFP

-56-

RNAs was translated and the resulting fusion proteins were assayed for retention of IL-2 bioactivity and specific nucleic acid binding ability (Examples 5, 6, and 7). All three triple-domain GFPs were found to retain these activities.

5

Example 9

Construction of a Targeted Nucleic Acid (tNA)

The yeast transcriptional activator, GAL4, has specific affinity for several closely related 17bp double-stranded DNA sequences, and it has also been shown to bind consensus synthetic 17bp target sequences with a similar affinity as it does wild-type sequences (Giniger, et al., Cell 40:767-774, 1985; Bram and Kornberg, P.N.A.S. 82:43-47, 1985). Target vectors were made by ligating oligomers containing a consensus 17bp sequence (Webster, et al., Cell 52:169-178, 1988; Carey, et al., J. Mol. Biol. 209:423-432, 1989) into the unique SalI site in the backbone of pDC302CAT (Mosley et al., Cell 59:335-348, 1989; and Overell et al., J. Imm. Meth. 141:53-62, 1991), a plasmid which directs the expression of chloramphenicol acetyl transferase (CAT) in mammalian host cells. The following oligomers were used:

5' TCGACGGAGTACTGTCCTCCGC 3'

3' GCCTCATGACAGGAGGCGAGCT 5'.

The oligomer pair harbored an internal ScaI site (TCATGA) which was used to screen resulting transformants for presence of the oligomer. In addition, there were SalI-compatible over-hangs at both the 5' and 3' ends, only one of which could regenerate the SalI site upon ligation. This feature was incorporated to allow the release of oligomer multimers for plasmid characterization. The oligomer pair was annealed by boiling equal amounts of each in a moderate salt buffer, then slow cooling the reaction. The annealed oligomers were then kinased with T4 polynucleotide kinase (Boehringer Mannheim) and ligated to pDC302CAT which had been linearized with SalI. The constructs were electroporated into DH10B *E. coli* and resulting colonies were screened for the presence of the target oligomer, or multimers thereof, by hybridization to a [³²P]-labeled target oligomer probe. Colonies harboring the target oligomer were further characterized for number of copies by restriction enzyme analysis and sequencing.

-57-

Example 10Ability of GDFPs to Bind to IL-2 Receptor-Bearing CTLL

GAL4/IL-2 and IL-2/GAL4 GDFPs from *in vitro* translations (Example 5) were further demonstrated to bind CTLL (see Example 7) via the following assay. CTLL were incubated in IL-2-free medium for 2 hours or longer. [³⁵S]-labeled GAL4/IL-2, IL-2/GAL4, IL-2, and GAL4 (Example 5) were incubated with the CTLL for 1 hour at 4 degrees C in a binding medium containing 25mg/ml BSA and 2mg/ml Na-azide in RPMI-1640 buffered with 20mM HEPES. The binding medium was adjusted to a final pH of 7.2 prior to use. After binding, the cells were washed three times in ice cold PBS, and the final cell pellet was resuspended in a Tris buffer containing 150mM NaCl, 5mM EDTA, 0.02% Na-azide, and 0.5% Triton X-100 to gently lyse the cells. The lysate was spun briefly, and the supernatant was electrophoresed through a 4-20% gradient polyacrylamide gel. Figure 9 shows labeled protein present in the CTLL lysate and, therefore, associated with the CTLL. Lane 1 shows a molecular weight standard. Lane 2 shows the human IL-2 protein as present in the unreacted reticulocyte lysate (Example 5), and lane 3 is the CTLL lysate after binding to IL-2. Lane 4 shows the GAL4/IL-2 GDFP as present in the unreacted reticulocyte lysate, and lane 5 is the CTLL lysate after binding to the GAL4/IL-2 GDFP. Lane 6 shows the IL-2/GAL4 GDFP as present in the unreacted reticulocyte lysate, and lane 7 is the CTLL lysate after binding to the IL-2/GAL4 GDFP. Lane 8 shows GAL4 as present in the unreacted reticulocyte lysate, and lane 9 is the CTLL lysate after binding to GAL4. The GAL4/IL-2 and IL-2/GAL4 GDFPs and IL-2 thus bind specifically to CTLL while GAL4 does not. This demonstrates that CTLL-specific binding of the GDFPs is mediated by the IL-2 domain and not by the GAL4 domain.

Example 11Ability of GAL4/IL-2 and IL-2/GAL4 GDFPs to Mediate Binding of a Target Oligomer to IL-2 Receptor-Bearing CTLL

GAL4/IL-2 and IL-2/GAL4 fusion proteins from *in vitro* translations (as in Example 5) were bound to [³²P]-dCTP-end-labeled GAL4 target oligomer as described in Example 6. The GDFP-tNA complex was bound to CTLL, as described in Example 10, for 1 hour at 4 degrees C in binding medium containing 25mg/ml BSA and 2mg/ml

-58-

Na-azide in RPMI-1640 buffered with 20mM HEPES. The binding medium was adjusted to a final pH of 7.2 prior to use. The cell-bound GDFP-tNA complex was separated from free GDFP-tNA by centrifugation of the binding mixture through a phthalate oil layer (Dower, et al., J. Exp. Med. 162:501-515, 1985). Cell-associated counts were quantified by scintillation counting. Figure 10 shows counts of labeled oligomer associated with CTLL as mediated by the GAL4/IL-2 GDFP, the IL-2/GAL4 GDFP, GAL4, and a negative control reticulocyte lysate designated "Bg." The binding assay demonstrates the ability of both GAL4/IL-2 and IL-2/GAL4 GDFPs to mediate binding of the oligomer tNA to CTLL.

Example 12

Ability of the GAL4/IL-2 GDFP to Mediate Binding of a Target Plasmid to IL-2

Receptor-Bearing CTLL

The GAL4/IL-2 GDFP (Example 5) was bound to the target plasmid using binding conditions described in Example 6. The plasmid contained eight copies of the GAL4 17-mer target oligomer, as described in Example 8. The GDFP-tNA complex was bound to CTLL for 1 hour at 4 degrees C in binding medium as described in Example 10. CTLL were then washed three times in ice cold PBS, and the final cell pellet was resuspended in a Tris buffer containing 150mM NaCl, 5mM EDTA, 0.02% Na-azide, and 0.5% Triton X-100 to gently lyse the cells. The cell lysate was spun briefly, the supernatant was brought to 0.4N NaOH, and the sample was denatured at 60 degrees C for 1 hour. The sample was then applied via slot-blot onto GeneScreen Plus membrane (NEN). The blot was screened for the presence of the target plasmid by hybridization to a [³²P]-labeled CAT probe. The membrane was washed, and the signal from cell associated plasmid was quantified by a phosphorimager (Molecular Dynamics). Figure 11 shows association of plasmid to CTLL mediated by either the GAL4/IL-2 GDFP or a negative control reticulocyte lysate designated "Bg." The binding assay showed the ability of the GAL4/IL-2 GDFP to mediate binding of plasmid tNA to CTLL.

Utility

The gene delivery fusion proteins of the present invention are useful in creating non-viral gene delivery systems for delivering a polynucleotide to a target cell.

-59-

Claims

1. A fusion protein useful in delivering a targeted nucleic acid to a target cell,
5 comprising a gene delivery fusion protein (GDFF), said GDFF comprising a nucleic acid binding domain (NBD) that binds to the targeted nucleic acid, fused to a gene delivery domain (GDD) that mediates delivery of the targeted nucleic acid to the target cell.
- 10 2. A fusion protein according to claim 1, wherein the targeted nucleic acid is a double-stranded nucleic acid.
3. A fusion protein according to claim 1, wherein the targeted nucleic acid is a single-stranded nucleic acid.
- 15 4. A fusion protein according to claim 1, wherein the targeted nucleic acid is DNA or an analog thereof.
5. A fusion protein according to claim 1, wherein the targeted nucleic acid is
20 RNA or an analog thereof.
6. A fusion protein according to claim 1, wherein the targeted nucleic acid is in the form of a recombinant expression vector comprising a nucleotide sequence to be expressed in the target cell.
- 25 7. A fusion protein according to claim 6, wherein the nucleotide sequence to be expressed is a nucleotide sequence that is not normally expressed in the target cell.
8. A fusion protein according to claim 6, wherein the nucleotide sequence to be
30 expressed is an antisense copy of a nucleotide sequence present in the target cell.

-60-

9. A fusion protein according to claim 1, wherein said GDFP further comprises a flexible polypeptide linker sequence ("flexon") between said nucleic acid binding domain and said gene delivery domain or within one of said domains.

5 10. A fusion protein according to claim 1, wherein said NBD comprises a nucleic acid binding component of a sequence-specific nucleic acid binding protein.

 11. A fusion protein according to claim 1, wherein said NBD comprises a nucleic acid binding component of a sequence-non-specific nucleic acid binding protein.
10

 12. A fusion protein according to claim 1, wherein said NBD comprises a multiplicity of nucleic acid binding (NB) components that bind one or more targeted nucleic acids.

15 13. A fusion protein according to claim 12, wherein said NBD comprises at least two NB components having differing binding specificities.

 14. A fusion protein according to claim 12, wherein the NBD comprises a first NB component capable of binding to a specific cognate recognition sequence present in
20 the targeted nucleic acid and a second NB component capable of binding non-specifically to the targeted nucleic acid.

 15. A fusion protein according to claim 1, wherein the NBD further comprises a component capable of mediating condensation and/or charge neutralization of the
25 targeted nucleic acid.

 16. A fusion protein according to claim 1, wherein said gene delivery domain (GDD) comprises one or more components that facilitate delivery of a targeted nucleic acid to a target cell.
30

 17. A fusion protein according to claim 16, wherein said components that facilitate delivery of a targeted nucleic acid to a target cell are selected from the group

-61-

consisting of a binding/targeting component, a membrane-disrupting component, a transport/localization component and a replicon integration component.

18. A fusion protein according to claim 16, wherein said GDD comprises two
5 or more components that facilitate delivery of a targeted nucleic acid to a target cell, said components selected from the group consisting of a binding/targeting component, a membrane-disrupting component, a transport/localization component and a replicon integration component.

19. A fusion protein according to claim 16, wherein said GDD comprises a
10 binding/targeting component.

20. A fusion protein according to claim 16, wherein said GDD comprises a
15 membrane disrupting component.

21. A fusion protein according to claim 16, wherein said GDD comprises a
transport/localization component.

22. A fusion protein according to claim 16, wherein said GDD comprises a
20 replicon integration component.

23. A fusion protein according to claim 22, wherein said replicon integration
component is an integrase enzyme or a derivative thereof that retains integrase activity.

24. A macromolecular complex useful in delivering a targeted nucleic acid to a
25 target cell, comprising a gene delivery fusion protein (GDFP) of claim 1 in association with a targeted nucleic acid.

25. A macromolecular complex according to claim 24, wherein said GDFP
30 comprises a replicon integration component.

-62-

26. A macromolecular complex according to claim 25, wherein said replicon integration component comprises a recombinase enzyme or a derivative thereof that retains recombinase activity, and wherein the targeted nucleic acid comprises NBD cognate recognition sequences in proximity to terminal recombinase recognition sequences.

27. A macromolecular complex according to claim 26, wherein said recombinase is an integrase enzyme or a derivative thereof that retains integrase activity, and wherein the targeted nucleic acid comprises NBD cognate recognition sequences in proximity to terminal integrase recognition sequences.

28. A recombinant polynucleotide useful for preparing a gene delivery fusion protein, said polynucleotide comprising a coding sequence that encodes a GDFP of claim 1.

29. The recombinant polynucleotide of claim 28, wherein said polynucleotide is in the form of an expression vector comprising a transcriptional control region operably linked to said coding sequence.

30. A cell useful in preparing a gene delivery fusion protein, said cell containing an expression vector of claim 29.

31. A method of using a recombinant polynucleotide of claim 29 to produce a GDFP, said method comprising the steps of causing the recombinant polynucleotide to be transcribed and translated, and recovering a GDFP.

32. A method of using a GDFP of claim 1 to deliver said targeted nucleic acid to a target cell, the method comprising the steps of contacting the GDFP with the targeted nucleic acid to produce a GDFP/nucleic acid complex and contacting said GDFP/nucleic acid complex with the target cell.

33. A cell produced by the method of claim 32 and progeny thereof.

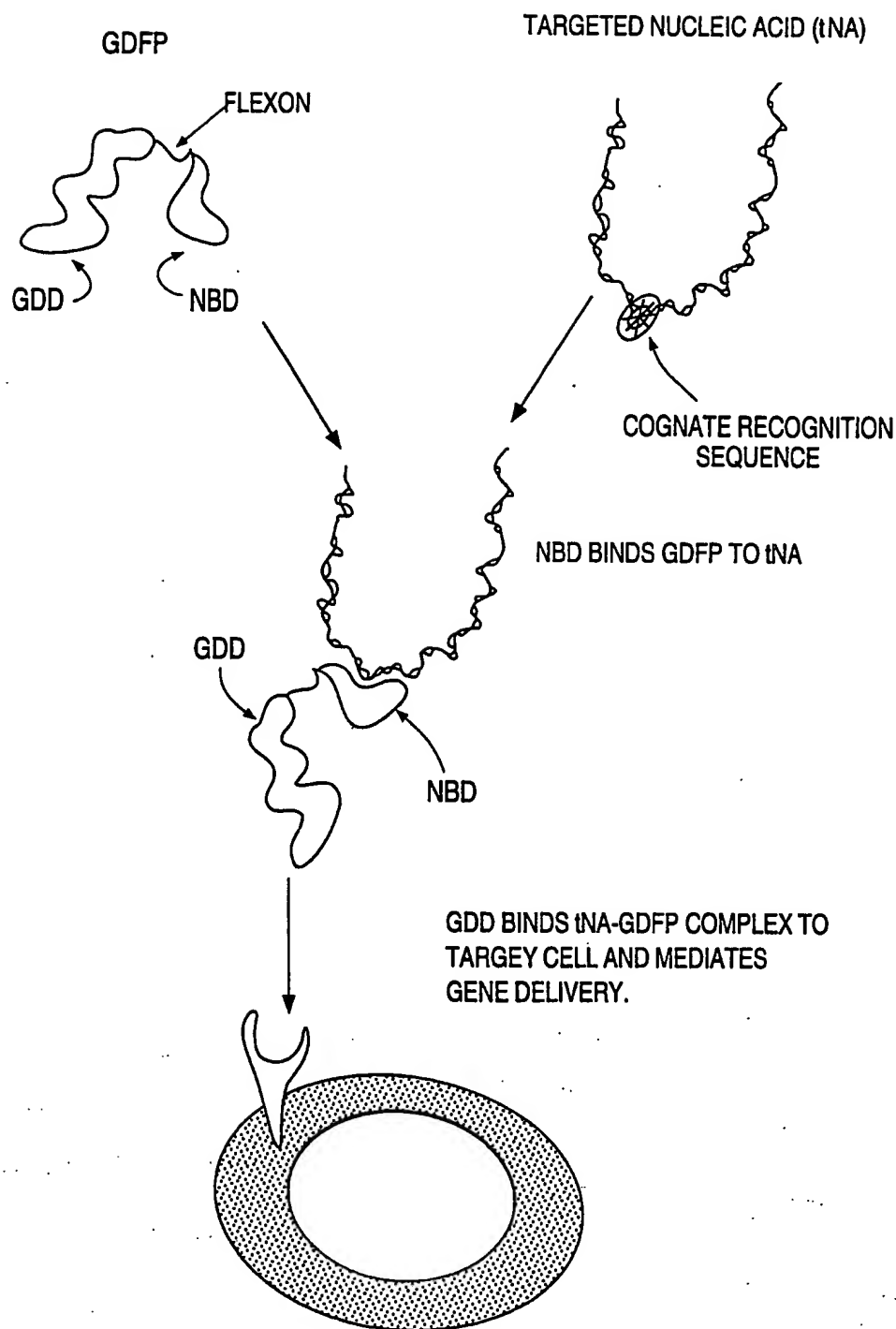
-63-

34. The cell of claim 33 wherein the targeted nucleic acid is expressed in the cell as an RNA molecule selected from the group consisting of an RNA transcript, an antisense RNA, an RNA decoy and a ribozyme.

5 35. A method of using a GDFP of claim 23 to deliver said targeted nucleic acid to a target cell, the method comprising the steps of contacting the GDFP with the targeted nucleic acid to produce a GDFP/nucleic acid complex and contacting said GDFP/nucleic acid complex with the target cell.

10 36. A cell produced by the method of claim 35 and progeny thereof, said cell comprising an integrated copy of said targeted nucleic acid.

1/12

**FIG. 1**

SUBSTITUTE SHEET (RULE 26)

2/12

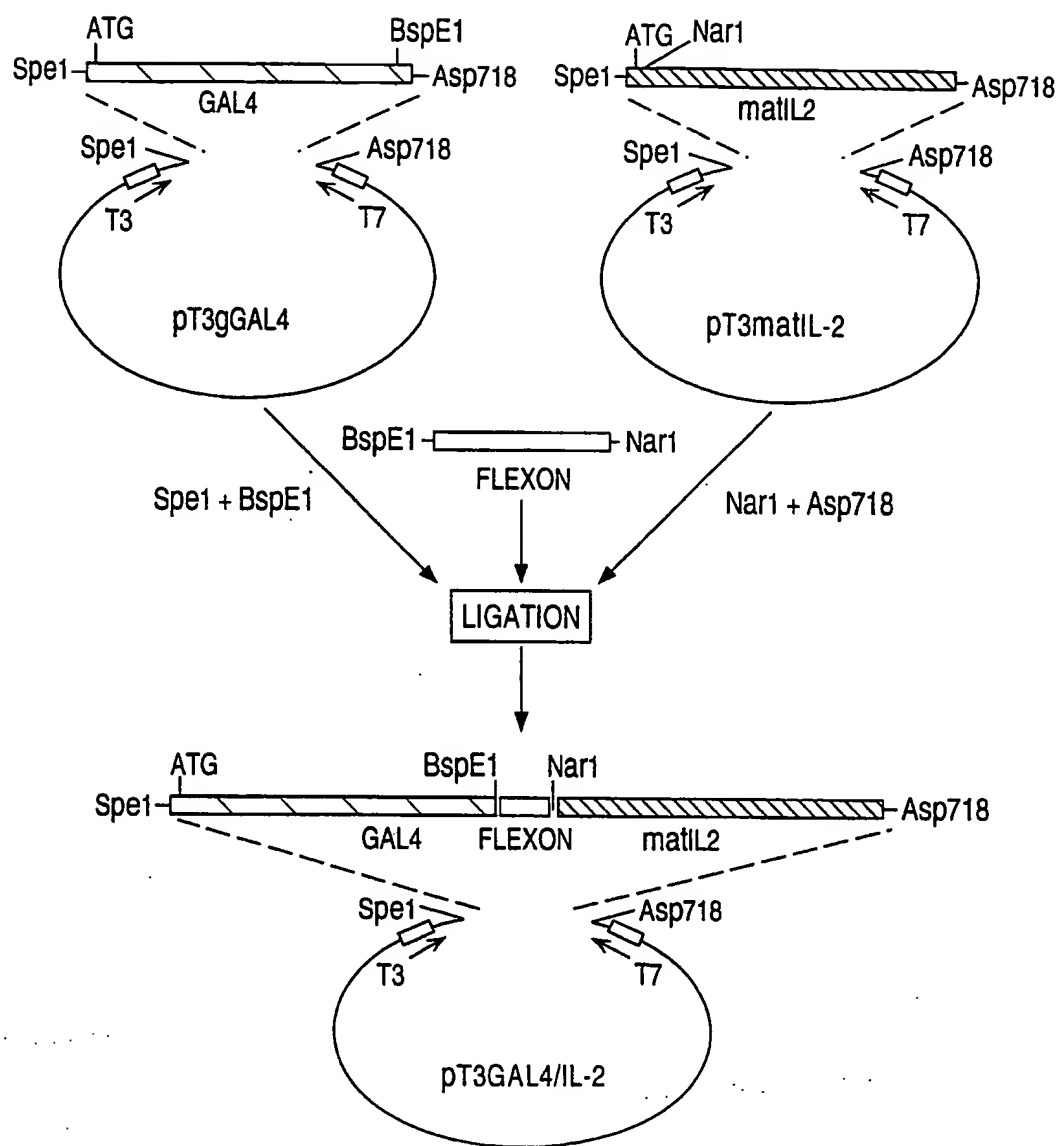


FIG. 2A

3/12

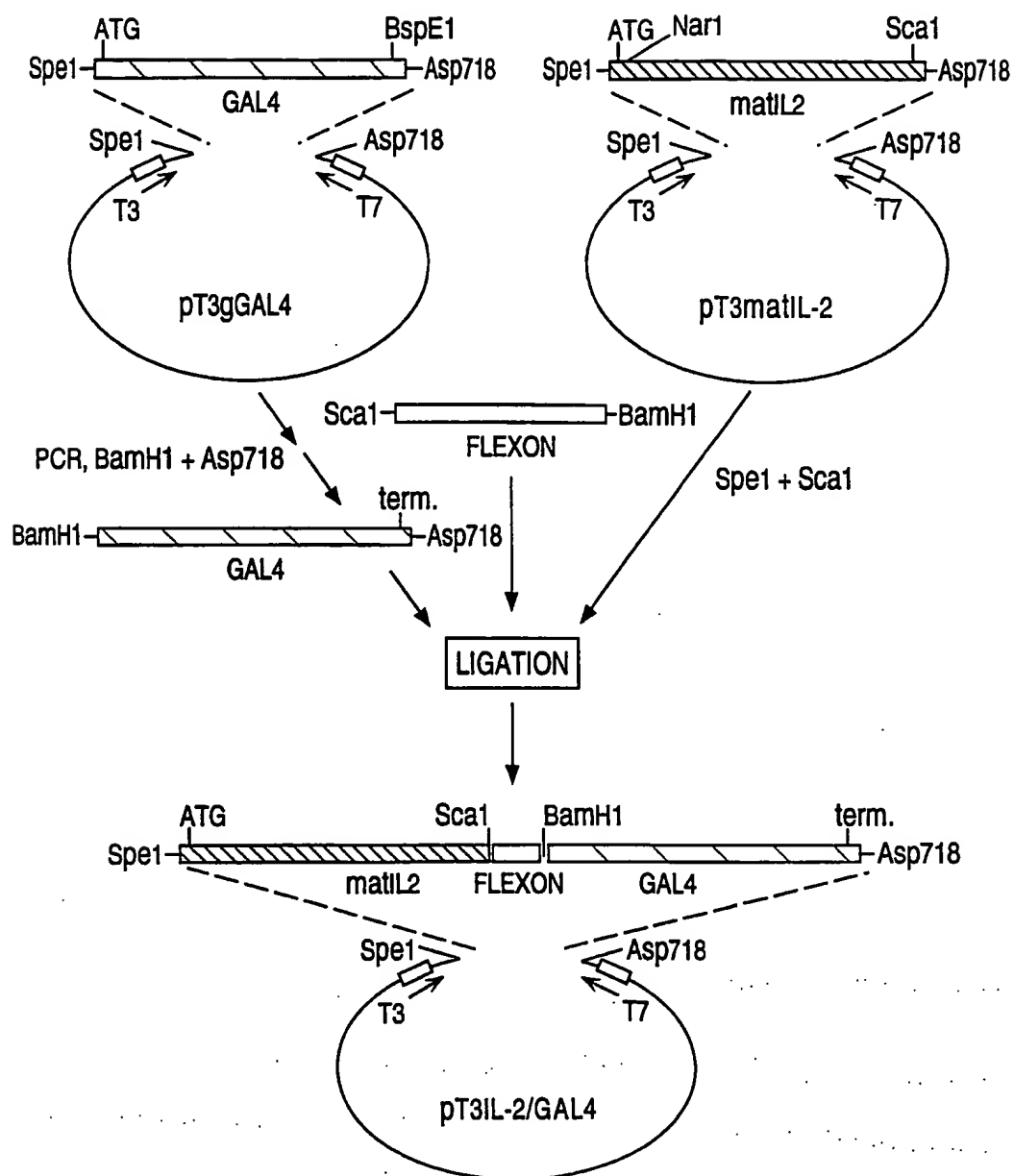


FIG. 2B

4/12

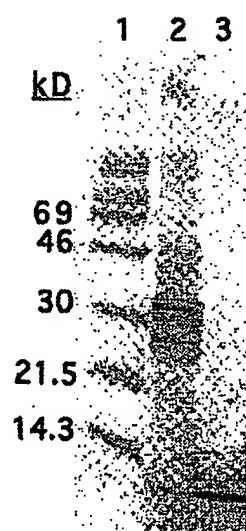


FIG. 3

5/12

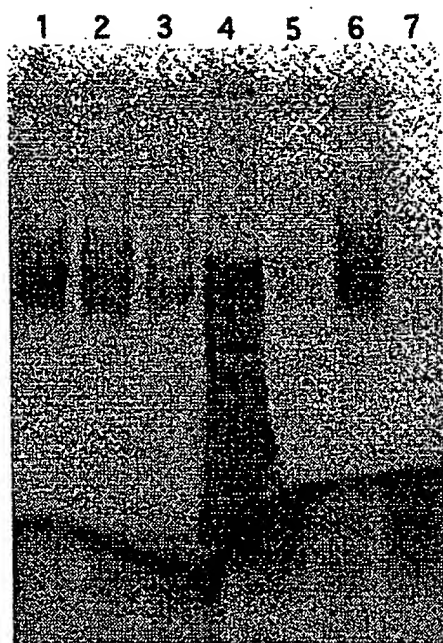
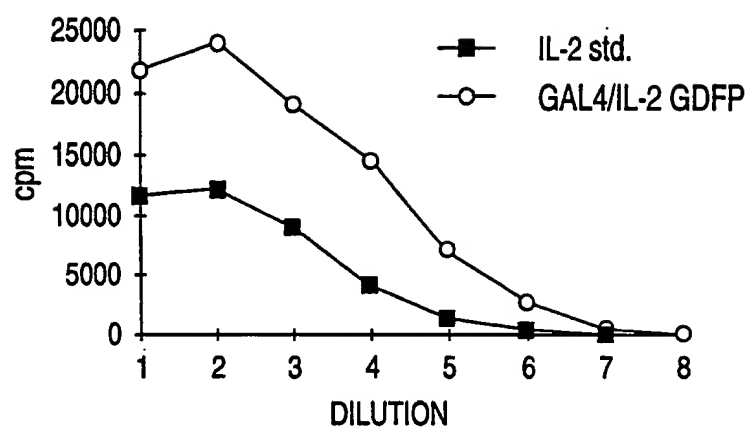


FIG. 4

6/12

**FIG. 5**

7/12

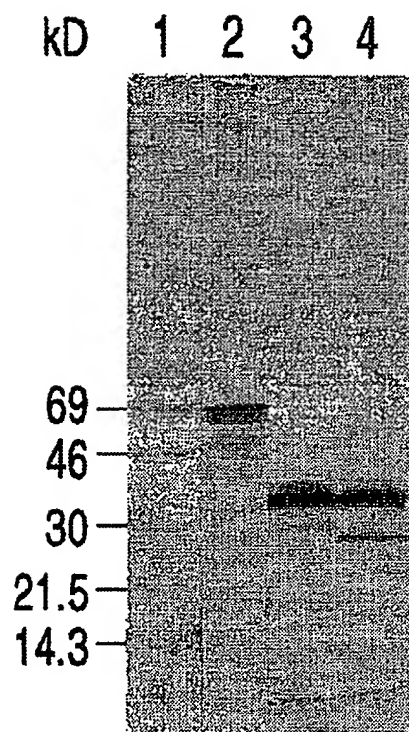


FIG. 6

8/12

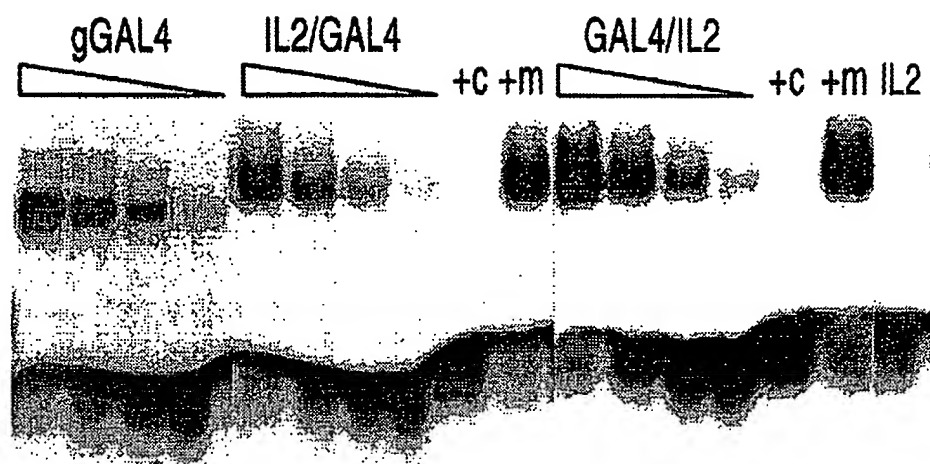


FIG. 7

9/12

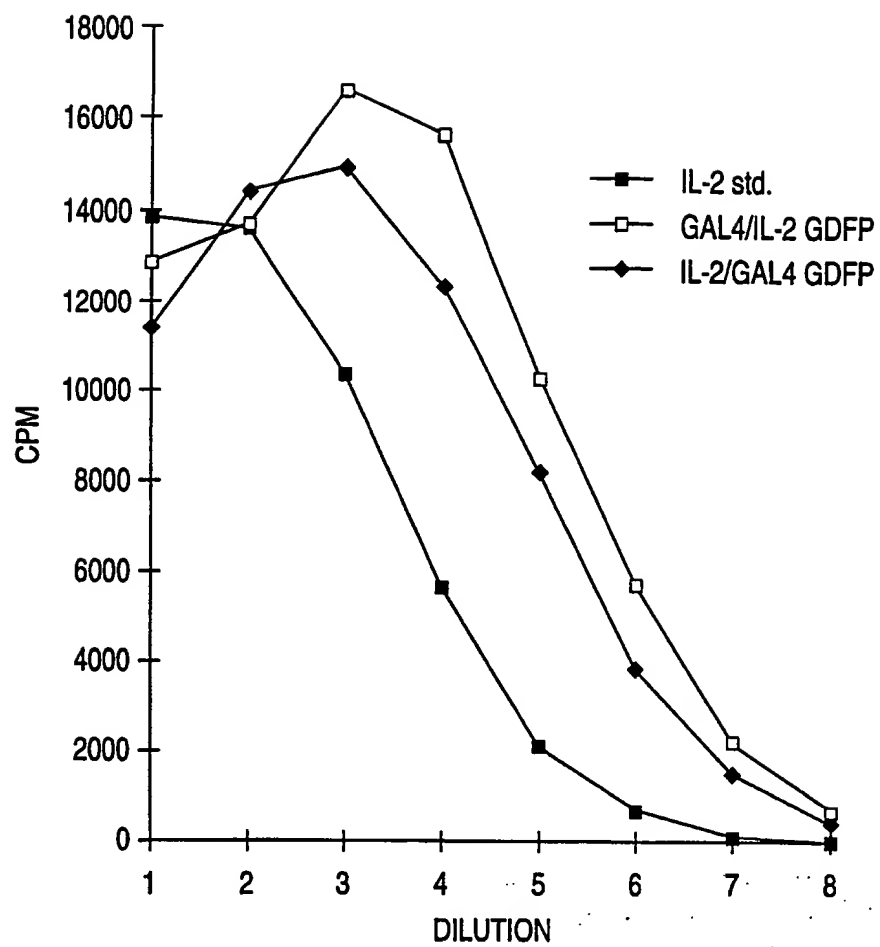


FIG. 8

10/12

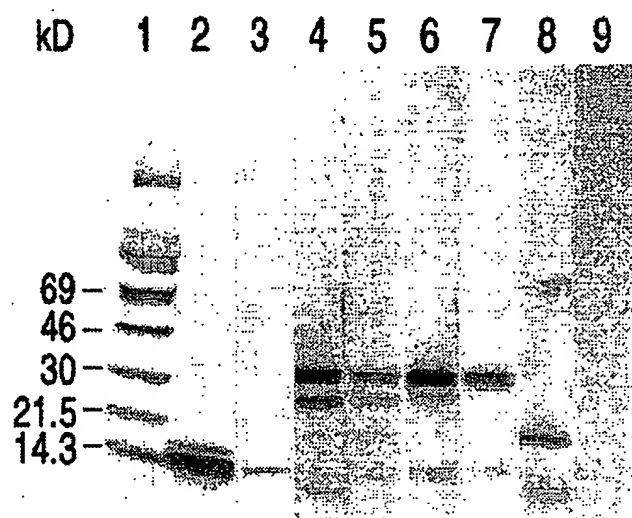
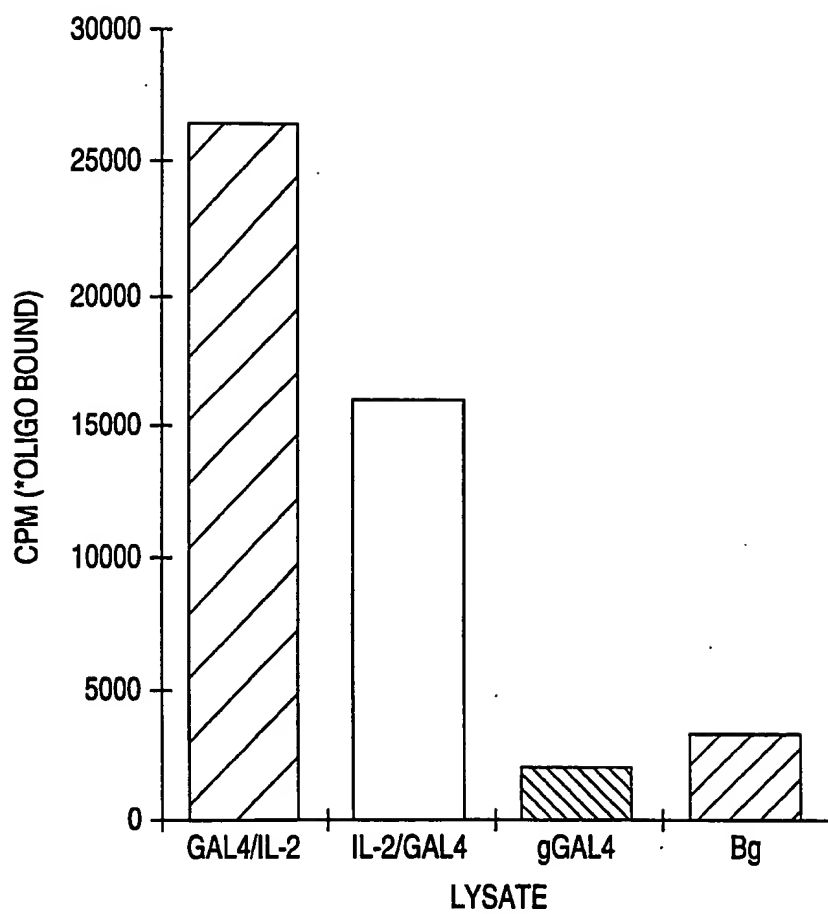
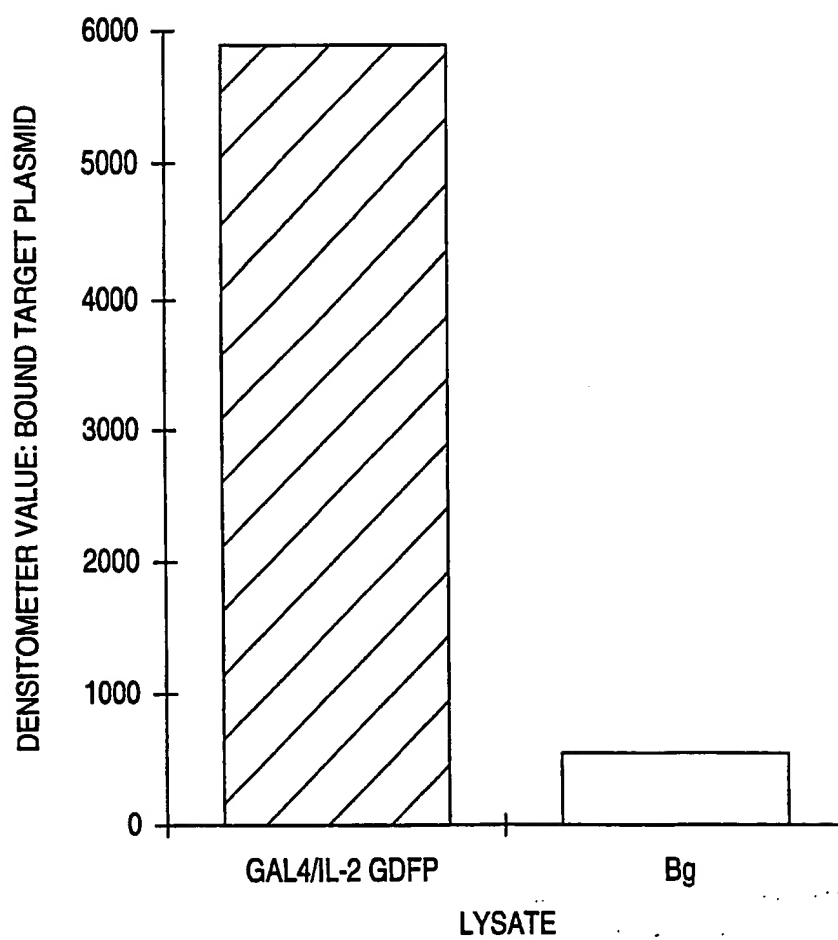


FIG. 9

11/12

**FIG. 10**

12/12

**FIG. 11**

INTERNATIONAL SEARCH REPORT

Internal Application No

PCT/US 95/04738

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/87 C12N15/62 C07K14/395 C07K14/55 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N A61K C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,94 04696 (MILES INC.) 3 March 1994 see the whole document ---	1-21,24, 28-36
X	WO,A,93 22443 (SRI INTERNATIONAL) 11 November 1993 see page 11, line 16 - page 12, line 9 ---	1-8, 10-18, 22-36
X	J VIROL, JUL 1993, 67 (7) P3777-85, UNITED STATES, COTTEN M ET AL 'Chicken adenovirus (CELO virus) particles augment receptor-mediated DNA delivery to mammalian cells and yield exceptional levels of stable transformants.' see the whole document ---	1-8, 10-18, 20,24, 28-36
-/--		



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

2

Date of the actual completion of the international search

9 August 1995

Date of mailing of the international search report

23.08.95

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Gurdjian, D

INTERNATIONAL SEARCH REPORT

Internat'l Application No
PCT/US 95/04738

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>PROC NATL ACAD SCI U S A, SEP 1 1992, 89 (17) P7934-8, UNITED STATES, WAGNER E ET AL 'Influenza virus hemagglutinin HA-2 N-terminal fusogenic peptides augment gene transfer by transferrin-polylysine-DNA complexes: toward a synthetic virus-like gene-transfer vehicle.' see the whole document -----</p>	<p>1-8, 10-18, 20,24, 28-36</p>

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 95/04738

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9404696	03-03-94	AU-B-	5088593	15-03-94
		CA-A-	2143308	03-03-94
		EP-A-	0658210	21-06-95
		FI-A-	950866	24-04-95
		NO-A-	950726	18-04-95

WO-A-9322443	11-11-93	AU-B-	4115693	29-11-93
		CA-A-	2118513	11-11-93
		JP-T-	7506252	13-07-95
